

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:52:49 ; Search time 116.78 Seconds
(without alignments)
34.125 Million cell updates/sec

Title: US-09-196-161d-4
Perfect score: 176
Sequence: 1 PFANNANRAGICVPCQINRVGSVTNAGDLATLAT 34

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP-archaea:*
2: SP-bacteria:*
3: SP-fungi:*
4: SP-human:*
5: SP-invertebrate:*
6: SP-mammal:*
7: SP-mhc:*
8: SP-organella:*
9: SP-phage:*
10: SP-plant:*
11: SP-rodent:*
12: SP-virus:*
13: SP-vertebrate:*
14: SP-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	176	100.0	394	5	Q27208
2	176	100.0	442	5	Q9XZG2
3	54	30.7	163	2	Q9L9W0
4	53.5	30.4	246	2	Q9L1A6
5	53	30.1	634	5	Q9VRW0
6	52	29.5	1817	11	Q54952
7	51	29.0	2249	5	Q9VCU9
8	50.5	28.7	384	4	Q9UBR1
9	50.5	28.7	387	4	Q9UR13
10	50.5	28.7	453	3	Q9PAR6
11	50	28.4	261	10	Q38828
12	49.5	28.1	115	2	Q50108
13	49.5	28.1	206	2	Q9RMP8
14	49.5	28.1	229	2	Q9RGR5
15	49.5	28.1	229	2	Q9LAI5
16	49.5	28.1	706	2	Q9KIK3
17	49	27.8	318	3	Q9Y7P4
18	49	27.8	357	13	Q91655
19	49	27.8	500	4	Q9Y5X9

20	49	27.8	517	5	Q9XYS3	Q9XYS3 dictyosteli
21	49	27.8	775	12	Q9WAK9	Q9WAK9 human rotav
22	49	27.8	1394	5	Q01787	Q01787 caenorhabdi
23	48.5	27.6	213	10	Q9LMM7	Q9LMM7 oryza sativ
24	48.5	27.6	706	2	Q9JWZ2	Q9JWZ2 neisseria m
25	48	27.3	205	12	Q9JUD31	Q9JUD31 human immun
26	48	27.3	370	1	Q9YBH6	Q9YBH6 aerypyrum p
27	48	27.3	845	5	Q9W3L2	Q9W3L2 drosophila
28	47.5	27.0	243	5	Q9RXD8	Q9RXD8 agelenopsis
29	47.5	27.0	3110	4	Q9J022	Q9J022 homo sapien
30	47	26.7	126	12	Q78400	Q78400 human immun
31	47	26.7	240	2	Q32259	Q32259 bacillus su
32	47	26.7	444	5	Q9UAE6	Q9UAE6 toxoplasma
33	47	26.7	1860	5	Q9U129	Q9U129 leishmania
34	46.5	26.4	245	2	Q54677	Q54677 streptomyce
35	46.5	26.4	604	2	P72304	P72304 rhodococcus
36	46.5	26.4	650	5	Q9ML28	Q9ML28 clona intes
37	46.5	26.4	1372	5	P91526	P91526 caenorhabdi
38	46.5	26.4	3512	5	Q62524	Q62524 chironomus
39	46	26.1	66	10	Q9XFY3	Q9XFY3 chlorella p
40	46	26.1	92	12	Q9YMR0	Q9YMR0 lymantria d
41	46	26.1	118	12	Q78916	Q78916 human immun
42	46	26.1	170	12	Q77380	Q77380 human immun
43	46	26.1	170	12	Q77381	Q77381 human immun
44	46	26.1	170	12	Q77382	Q77382 human immun
45	46	26.1	170	12	Q80734	Q80734 human immun

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	394 AA.
ID	Q27208			
AC	Q27208			
DT	01-NOV-1996 (TRENBLREL. 01, Created)			
DT	01-NOV-1998 (TRENBLREL. 08, Last sequence update)			
DT	01-JUN-2000 (TRENBLREL. 14, Last annotation update)			
DE	IMMOBILIZATION ANTIGEN PRECURSOR (FRAGMENT).			
OS	Ichthyophthirius multifiliis.			
OC	Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;			
OC	Opisthokonta; Ichthyophthirius multifiliis.			
OX	NCBI_TaxID=5932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=GEORGIA;			
RC	MEDLINE=92335298; PubMed=1631132;			
RA	Clark T.G., McGraw R.A., Dickerson H.W.;			
RT	"Developmental expression of surface antigen genes in the parasitic			
RT	ciliate Ichthyophthirius multifiliis";			
RL	Proc. Natl. Acad. Sci. U.S.A. 89:6363-6367(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=GEORGIA;			
RC	MEDLINE=93020590; PubMed=1383510;			
RA	Lin T.L., Dickerson H.W.;			
RT	"Purification and partial characterization of immobilization antigens			
RT	from Ichthyophthirius multifiliis";			
RL	J. Protozool. 39:457-463(1992).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=GEORGIA;			
RC	Clark T.;			
RA	Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.			
RL	Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=GEORGIA;			
RC	Clark T.;			
RA	Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.			
RL	Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; M92907; AAC36158.1; -.			
KW	Signal.			
FT	NON_TER			
FT	SIGNAL			

```

SQ SEQUENCE 394 AA; 39495 MW; 3013C2B2BEFDB682 CRC64;

Query Match          100.0%; Score 176; DB 5; Length 394;
Best Local Similarity 100.0%; Pred. No. 1e-17;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 PFAANNAARGICVPCOINRVGSVTNAGDLATLAT 34
        |||||||
Db       47 PFAANNAARGICVPCOINRVGSVTNAGDLATLAT 80

RESULT 2
O9XZG2 PRELIMINARY; PRT; 442 AA.
AC O9XZG2;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE IMMOBILIZATION ANTIGEN PRECURSOR.
GN IAg48.
OS Ichthyophthirius multifiliis.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OX Opithyoglenina; Ichthyophthirius.
NX NCBI_TaxID=5932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G1;
RX MEDLINE=99196987; PubMed=10095108;
RA Clark T.G., Lin T.L., Jackwood D.A., Sherrill J., Lin Y.,
RA Dickerson H.W.;
RT "The gene for an abundant parasite coat protein predicts tandemly
RT repetitive metal binding domains.";
RL Gene 229:91-100(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=G1;
RA Gaetig J., Gao Y., Tishgarten T., Clark T.G., Dickerson H.W.;
RT "Surface display of a parasite antigen in the ciliate Tetrahymena
RT thermophila.";
RL Nat. Biotechnol. 0:0-0(1999).
DR EMBL: AF140273; AAD31283.1; .
KW signal.
FT SIGNAL.
FT CHAIN 1 20 POTENTIAL.
FT CHAIN 21 442 IMMOBILIZATION ANTIGEN.
SQ SEQUENCE 442 AA; 45025 MW; 52658F3F65D27AFA CRC64;

Query Match          100.0%; Score 176; DB 5; Length 442;
Best Local Similarity 100.0%; Pred. No. 1.e-17;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 PFAANNAARGICVPCOINRVGSVTNAGDLATLAT 34
        |||||||
Db       67 PFAANNAARGICVPCOINRVGSVTNAGDLATLAT 100

RESULT 3
O9L9W0 PRELIMINARY; PRT; 163 AA.
AC O9L9W0;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE PLIIN (FRAGMENT).
GN PLE.
OS Neisseria cinerea.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
NX NCBI_TaxID=483;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRL 33683;
RX MEDLINE=20112990; PubMed=10644494;

```

```

RA  Aho E.L. Keating A.M., McGillivray S.M.;
RT  "A comparative analysis of pilin genes from pathogenic and
RL  nonpathogenic neisseria species.";
DR  Microb. Pathog. 28:81-88(2000).
SQ  EMBL: AF158604; AAF43788.1; -.
FT  NON_TER 163 163
SQ  SEQUENCE 163 AA; 16732 MW; C2345EDFPFF18FBD CRC64;

Query Match
Best Local Similarity 30.7%; Score 54; DB 2; Length 163;
Matches 14; Conservative 1; Mismatches 8; Indels 2; Gaps 1;

OY 2 FAANNAARGICVPCOI--NRVGSVT 24
    | | | | | | | | | | | | | | | |
DB 64 FPAANNAAGIAAPTEIKGRYASVT 88

RESULT 4
O9LIA6 PRELIMINARY; PRT: 246 AA.
ID O9LIA6
AC O9LIA6;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DI 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DR 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE PUTATIVE THIAMINE BIOSYNTHESIS LIPOPROTEIN PRECURSOR.
GN SCJ24.11.
OC Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinash H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL157956; CAB76095.1; -.
KW Lipoprotein.
SQ SEQUENCE 246 AA; 25098 MW; 63F5DA13C7514BDD CRC64;

}

Query Match
Best Local Similarity 30.4%; Score 53.5; DB 2; Length 246;
Matches 14; Conservative 3; Mismatches 15; Indels 1; Gaps 1;

OY 2 FAANNAARGICVPCOINRGSVTNAGDLATLAT 34
    | | | | | | | | | | | | | | | |
DB 96 WAVERAARGIAACGASGV-SVNGGGDYQLLET 127

RESULT 5
O9VRW0 PRELIMINARY; PRT: 634 AA.
ID O9VRW0
AC O9VRW0;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DI 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DR 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE LANA.
GN LANA.
SQ Drosophila melanogaster (Fruit fly).

```



```

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yeaman M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer V., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernier B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Chertys J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fostel C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maltel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sidenklamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003563; AAF50672.1; -
DR HSSP; P02468; ITLE.
DR FLYBASE; FBgn0002526; Lana.
DR INTERPRO; IPR000561; -
DR INTERPRO; IPR001886; -
DR INTERPRO; IPR002049; -
DR PFAM; PF00053; laminin_EGF_7.
DR PFAM; PF00053; laminin_Nterm; 1.
DR PRINTS; PR00011; EGF_LAMININ.
DR PROSITE; PS00022; EGF_1; UNKNOWN_5.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 6.
SO SEQUENCE 634 AA; 70927 MW; 51FC7DB308E9449 CRC64;

```

```

AC 054952; p97951;
DT 01-JUN-1998 (TRENBLER. 06, Created)
DT 01-JUN-1998 (TRENBLER. 06, Last sequence update)
DT 01-JUN-2000 (TRENBLER. 14, Last annotation update)
DE BREAST CANCER TYPE 1 SUSCEPTIBILITY PROTEIN (BREAST AND OVARIAN CANCER
DE SUSCEPTIBILITY PROTEIN).
DE
GN BCR1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY;
RX MEDLINE=99111388; PubMed=9892727;
RA Bennett L.M., Brownlee H.A., Hagavik S., Wiseman R.W.;
RT "Sequence analysis of the rat bcr1 homolog and its promoter region."
RL Mamm. Genome 10:19-25(1999).
RN [2]
RP SEQUENCE OF 8-222 FROM N.A.
RC STRAIN-WISTAR-KYOTO; TISSUE=SPLEEN;
RX MEDLINE=96358532; PubMed=8761410;
RA Chen K.S., Shepel L.A., Haag J.D., Heil G.M., Gould M.N.;
RT "Cloning, genetic mapping and expression studies of the rat Bcr1
RT gene."
RL Carcinogenesis 17:1561-1566(1996).
CC - FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION, MAY PLAY A ROLE
CC IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).
CC - SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC - SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
DR EMBL; AF036760; AAC36493.1; -
DR EMBL; S82504; -; NOT_ANNOTATED_CDS.
DR EMBL; S82502; -; NOT_ANNOTATED_CDS.
DR EMBL; U60523; AAB40387.1; -
DR EMBL; S82500; AAB37501.1; -
DR INTERPRO; IPR001357; -
DR INTERPRO; IPR001841; -
DR INTERPRO; IPR002378; -
DR PFAM; PF00097; zf-C3HC4; 1.
DR PFAM; PF00533; BRCT; 2.
DR PRINTS; PR00493; BRSTCANCER1.
DR PROSITE; PS00518; ZINC_FINGER_C3HC4; 1;
RW ZINC-finger; 24 64
FT ZN-FING 497 503 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 38 38 Q->K (IN REF. 2).
FT CONFLICT 192 192 A->M (IN REF. 2).
FT SEQUENCE 1817 AA; 199876 MW; C0B4760F0E349A01 CRC64;

```

```

Query Match 30.1%; Score 53; DB 5; Length 634;
Best Local Similarity 40.0%; Pred. No. 14;
Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

```

```

OY 12 CVCPCQINRGVSNTNAGDLAT 31
   1 : : : : : : : : : : : :
Db 492 CKACECNKIGSITDCNVT 511

```

```

RESULT 6 PRELIMINARY; PRT; 1817 AA.
054952
ID 054952

```

```

Query Match 29.5%; Score 52; DB 11; Length 1817;
Best Local Similarity 43.5%; Pred. No. 60;
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

```

```

OY 10 GICVPCQINRGVSNTNAGDLATL 32
   1 : : : : : : : : : : : :
Db 917 GLPVCQEGKCAVTCADVSRL 939

```

```

RESULT 7 PRELIMINARY; PRT; 2249 AA.
09VCU9
AC 09VCU9;
DT 01-MAY-2000 (TRENBLER. 13, Created)
DT 01-MAY-2000 (TRENBLER. 13, Last sequence update)
DT 01-OCT-2000 (TRENBLER. 15, Last annotation update)
DE CG4792 PROTEIN.
DE
GN CG4792.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;

```

```

[1]
RN SEQUENCE FROM N.A.
RP STRAIN-BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
  Amanalides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
  George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
  Sultn G.G., Wortman J.R., Yeandle M.D., Zhang Q., Chen L.X.,
  Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
  Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
  Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
  Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
  Beeson K.Y., Bens P.V., Berman B.P., Bhandari D., Bolshakov S.,
  Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
  Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
  Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
  de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
  Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
  Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
  Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
  Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
  Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
  Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
  Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
  Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
  Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
  Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
  Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
  Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
  Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclik J.M.,
  Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
  Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
  Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
  Spier E., Spradling A.C., Stapleton M., Strong R., Stun E.,
  Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
  Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
  Williams S.M., Woodage T., Wolley K.C., Wu D., Yang S., Yao Q.A.,
  Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
  Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
  RA Glbbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
  RT "The genome sequence of Drosophila melanogaster.";
  RL Science 287:2185-2195(2000)
DR EMBL: AE003740; AAF56056.1; -
DR FlyBase: FBgn0039016; CG4792.
DR INTERPRO: IPR000999; -
DR INTERPRO: IPR001159; -
DR INTERPRO: IPR001410; -
DR INTERPRO: IPR001650; -
DR INTERPRO: IPR003100; -
DR PFM: PFM00035; dsrm; 1.
DR PFM: PFM00271; helicase_C; 1.
DR PFM: PFM00636; Ribonuclease_3; 2.
DR PFM: PFM02170; ZAP; 1.
DR PROSITE: PS00517; RNase_3.1; 1.
SO SEQUENCE 2249 AA; 253328 MW; D693F0432AC8033D CRC64;

Query Match 29.0%; Score 51; DB 5; Length 2249;
Best Local Similarity 31.0%; Pred. No. 1.le+02;
Matches 9; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

```

```

DE BETA-UREIDOPROPIONASE.
GN BUP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN SEQUENCE FROM N.A.
RP STRAIN-CAUCASIAN; TISSUE=LIVER;
RC Sakamoto T.;
RA "Beta-ureidopropionase mRNA.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=20011284; PubMed=10542323;
RA Vreken P., van Kullenburg A.B., Hamajima N., Meisner R.,
  van Lenthe H., Gohllich-Ratmann G., Assmann B.E., Wevers R.A.,
  van Gent A.H.;
  RT "cDNA cloning, genomic structure and chromosomal localization of the
  human BUP-1 gene encoding beta-ureidopropionase(1).";
  RL Biochim. Biophys. Acta 1447:251-257(1999).
DR EMBL: AB013885; BAA88634.1; -
DR EMBL: AF163312; AAF06735.1; -
DR INTERPRO: IPR003010; -
DR PFM: PFM00795; CN_hydrolase; 1.
KW Hydrolase.
SO SEQUENCE 384 AA; 43166 MW; 62B81982D2D63CC3 CRC64;

```

```

Query Match 28.7%; Score 50.5; DB 4; Length 384;
Best Local Similarity 52.2%; Pred. No. 20;
Matches 12; Conservative 1; Mismatches 9; Indels 1; Gaps 1;

OY 1 PFAANNA-RGICVPCQINRGVS 22
DB 269 PFAANNAIAHCTCAINRGVT 291

RESULT 9
AC O9UBR1 PRELIMINARY; PRT; 387 AA.
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE BETA-UREIDOPROPIONASE.
GN BUP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Vreken P., van Kullenburg A.B.P., Hamajima N., Meisner R.,
  van Gent A.H.;
  RT "cDNA cloning, genomic structure and chromosomal localization of the
  human BUP-1 gene encoding beta-ureidopropionase.";
  RL Biochim. Biophys. Acta 1447:251-257(1999).
DR EMBL: AF169559; AAF06739.1; -
DR EMBL: AF169550; AAF06739.1; JOINED.
DR EMBL: AF169551; AAF06739.1; JOINED.
DR EMBL: AF169552; AAF06739.1; JOINED.
DR EMBL: AF169553; AAF06739.1; JOINED.
DR EMBL: AF169554; AAF06739.1; JOINED.
DR EMBL: AF169555; AAF06739.1; JOINED.
DR EMBL: AF169556; AAF06739.1; JOINED.
DR EMBL: AF169557; AAF06739.1; JOINED.
DR EMBL: AF169558; AAF06739.1; JOINED.
DR INTERPRO: IPR003010; -
DR PFM: PFM00795; CN_hydrolase; 1.
SO SEQUENCE 387 AA; 43522 MW; 7DFC31F0BF0C4C34 CRC64;

```


OX	NCBI_TaxID=33046;
RN	[1]
SEQUENCE	FROM N.A.
RP	SPRAIN=MM9.
RC	
RA	Birtles R.U., Hazel S., Bown K., Raoult D., Begon M., B
RT	"Subtyping of uncultured Bartonella using sequence com
KT	16S/23S rRNA intergenic spacer regions amplified direct
FT	infected bloods.";
RL	Submitted (Oct.1999) to the EMBL/Genbank/DBJ databases
DR	EMBL; AF1915102; AAF05721.1; -.
DR	HSSP; Q53554; 1A78.
DR	INTERPRO; IPR002020; -.
DR	PFAM; PF00285; citrate_synth; 1.
DR	PRINTS; PR00143; CITRNSHASE.
DR	PROSITE; PS00480; CITRATE_SYNTHASE; 1.
FT	NON_TER
FT	1
FT	1
FT	206
FT	206
SEQUENCE	206 AA; 23124 MW; 64F35E52CD39FAD4 CRC64;

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:47:59 ; Search time 70.34 Seconds

(without alignments)
101.359 Million cell updates/sec

Title: US-09-196-161d-1

Perfect score: 558
Sequence: 1 GAAGGEANGNPFANNMARNAR.....PGEAPGVVFAGAAAGV 105

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : PIR_66:*

1: PIR1:
2: PIR2:
3: PIR3:
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	558	100.0	395	2 A46031	Immunoblotting sur
2	93	16.7	677	2 C42125	trophozoite cystei
3	84.5	15.1	1737	2 T00209	MCF8 protein-hu
4	78	14.0	596	2 A45664	variant-specific s
5	77.5	13.9	573	2 T10053	laminin alpha 5 ch
6	77.5	13.8	1372	2 D83130	probable iron-sulf
7	77	13.8	1372	2 T25933	hypothetical prote
8	77	13.8	3712	2 S18253	laminin alpha-1 ch
9	75.5	13.5	713	2 A35502	major surface-labe
10	75	13.4	1111	2 B44018	laminin B2t chain
11	75	13.4	1193	2 A44018	laminin B2t chain
12	75	13.4	3084	1 M44018	laminin alpha-1 ch
13	72.5	13.0	398	1 S24802	polyferredoxin 6x2
14	72.5	13.0	2225	1 S18323	hypothetical prote
15	72	12.9	109	2 S18323	thyroglobulin - bu
16	71	12.7	1713	2 A55347	adhesive ligand ep
17	70.5	12.6	289	2 T25682	hypothetical prote
18	70.5	12.6	1607	1 M44018	laminin gamma-1 ch
19	70	12.5	1607	1 M44018	E-selectin - bovin
20	69.5	12.5	2824	2 T22759	hypothetical prote
21	69	12.4	1557	2 T22811	hypothetical prote
22	68.5	12.3	484	2 S77602	ccog protein - Par
23	68.5	12.3	962	2 JC5571	subtilisin-like pr
24	68.5	12.3	969	1 A39490	subtilisin-like pr
25	68.5	12.3	975	2 JC5570	subtilisin-like pr
26	68.5	12.3	1280	2 A39117	170k lectin precu
27	68.5	12.3	1895	2 T15881	hypothetical prote
28	68.5	12.3	2014	2 T21560	hypothetical prote
29	68	12.2	738	2 S40992	hypothetical prote

30	67.5	12.1	932	2 I52527	PACEA - mouse (fr
31	67	12.0	294	2 T21668	hypothetical prote
32	67	12.0	314	2 I37383	FAS soluble protei
33	67	12.0	335	2 A40036	apoptosis-mediatin
34	67	12.0	439	2 A36385	surface antigen se
35	67	12.0	1639	1 M44018	laminin gamma-1 ch
36	67	12.0	1827	2 T34288	hypothetical prote
37	67	12.0	3075	2 S14458	laminin alpha-1 ch
38	66.5	11.9	225	2 S48780	LI protein - human
39	66.5	11.9	309	2 F83044	nitrate-inducible
40	66.5	11.9	495	2 G82371	FIXG-related prote
41	66.5	11.9	820	2 H81106	ATP-dependent prot
42	66.5	11.9	846	2 A30889	Integrin beta chai
43	66	11.8	303	2 T19289	hypothetical prote
44	66	11.8	557	2 A48434	variant-specific s
45	66	11.8	1084	2 T08583	cellulose synthase

ALIGNMENTS

RESULT 1

A46031 Immunoblotting surface I-antigen precursor - Ichthyophthirius multifiliis (fragment)

C:Species: Ichthyophthirius multifiliis

C>Date: 29-Sep-1999 #sequence_revision 29-Sep-1999 #text_change 07-Dec-1999

C:Accession: A46031

R:Clark, T.G.; McGraw, R.A.; Dickerson, H.W.

A:Title: Developmental expression of surface antigen genes in the parasitic ciliate I

A:Reference number: A46031; MUID:92335298

A:Accession: A46031

A:Molecule type: mRNA; protein

A:Residues: 1-395 <CLAF>

A:Cross-references: GB:M92907; NID:93628568; PIDN:AAC36158.1; PID:93628569

A>Note: the authors translated the codon UUG for residue 330 as Ile

A:Note: sequence extracted from NCBI backbone (NCBIN:108734, NCBIPI:108735); the sequ

C:Genetics:

A:Genetic code: SGC5

C:Keywords: glycoprotein; surface antigen

F:2-395/Product: Immunoblotting surface I-antigen #status experimental <MA>

F:156,191,245,281/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match Best Local Similarity 100.0%; Score 558; DB 2; Length 395;

Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGGEANGNPFANNMARNARICVPCQINRVGSVTNAGDLATLTAQCSTGCTGALDDG 60

DB 37 GAAGGEANGNPFANNMARNARICVPCQINRVGSVTNAGDLATLTAQCSTGCTGALDDG 96

QY 61 VTVDENSAOCVCKPNEFYNGSPGGEAPGVVFAGAAAGV 105

DB 97 VTVDENSAOCVCKPNEFYNGSPGGEAPGVVFAGAAAGV 141

RESULT 2

C42125 trophozoite cysteine-rich surface antigen 72 - Giardia lamblia (fragment)

N:Alternate names: CRP72

C:Species: Giardia lamblia

C>Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 05-Dec-1998

C:Accession: C42125

R:Adam, R.D.; Yang, Y.M.; Nash, T.E.

Mol. Cell. Biol. 12, 1194-1201, 1992

A:Title: The cysteine-rich protein gene family of Giardia lamblia: loss of the CRP170

A:Reference number: A42125; MUID:92186850

A:Accession: C42125

A:Molecule type: DNA

A:Residues: 1-677 <ADA>

A:Cross-references: GB:M83934; NID:g159123

A:Experimental source: trophozoites


```

OY      2 AAGGNGNMGNPFAANNAAGI--CYPCQINRVGSVTNAGDLATLANQCSTGCCPTGTLD    59
          |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      350 AARGESRGRARRKKGSDDPRAGGLDCTDCQ-----QQVVCPTGTDIRD    391
          |::|::|::|
OY      60 GYTDFVFDRSAACVKC    75
          |::|::|
Db      392 GL-----QINACTGC    400

RESULT      7
T5933
hypothetical protein W02C12.1 - Caenorhabditis elegans
C.Species: Caenorhabditis elegans
C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C.Accession: T25933
R.Murray, J., Wohlmann, P.
submitted to the EMBL Data Library, December 1996
A.Description: The sequence of C. elegans cosmid W02C12.
A.Reference number: Z20112
A.Accession: T25933
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-1372 <MOR>
A.Cross-references: EMBL,U80815; PIDN:AAB37995.1; GSPDB:GN00022; CESP:W02C12.1
A.Experimental source: strain Bristol N2; clone W02C12
C.Genetics:
A.Gene: CESP:W02C12.1
A.Map position: 4
A.Introns: 29/1; 66/1; 774/2; 823/2; 1046/1; 1108/2; 1298/1

```

```

Query Match Similarity 13.8%; Score 77; DB 2; Length 1372;
Best Local Similarity 21.2%; Pred. No. 13;
Matches 28; Conservative 14; Mismatches 50; Indels 40; Gaps 5;

OY 3 AAGGANGNOPFANNAAGICVPCOINRVGS-----VTNAGDLTLATQCSST 49
      :: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1036 SDEEDRDECPDSEQJLSASGVQCPQDICTGRSRGENKKVCACPGCTTATMTREQCNT 1095
      :: | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 50 Q-----CPTGTALDDGVTVDFDRSAQCVCYKCRNFYNGSGPQGEAPG 92
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1096 FCKRPGQFLVETKNCQPCPRGTFQN-----EQESTCKLCAPD--HTTAA-P-GATAE 1145
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 93 VQVFAGCAAAAG 104
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1146 SOCFSTNOCATG 1157

RESULT 8
518253
Laminin alpha-1 chain precursor - fruit fly (Drosophila melanogaster)
c:Species: Drosophila melanogaster
c:Date: 16-Sep-1992 #sequence_revision 24-Jul-1997 #text_change 10-Dec-1999
c:Accession: S28339; S18253
R:Kusche-Gullberg, M.; Garritson, K.; Mackrell, A.J.; Fessler, L.I.; Fessler, J.H.
EMBO J. 11, 4519-4527, 1992
A:Title: Laminin A chain: expression during Drosophila development and genomic sequence
A:Reference number: S28339; MUID:93049203
A:Accession: S28339
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-3712 <KUS>
A:Cross-references: GB:M96388; NID:g157799; PIDN:AAA28662.1; PID:g157800
R:Garritson, K.; Mackrell, A.J.; Fessler, J.H.
R: Biol. Chem. 266, 22899-22904, 1991
A:Title: Drosophila laminin A chain sequence, interspecies comparison, and domain structure
A:Reference number: S18253; MUID:92078147
A:Accession: S18253
A:Molecule type: mRNA
A:Residues: 1762-3712 <GAR>
A:Cross-references: EMBL:M75882; NID:g157797; PIDN:AAA28661.1; PID:g157798
;Genetics:

```

A:Gene: FlyBase:lanA
A:Cross-references: FlyBase:FBgn0002526
C:Superfamily: laminin alpha-1 chain, laminin G repeat homology; laminin-type EGF-like C:Keywords: basement membrane; cell binding; coll; disulfide bond; extracellular
E:273-30/Domain: laminin-type EGF-like homology <LEG>
F:333-400/Domain: laminin-type EGF-like homology <LEG1>
F:541-584/Domain: laminin-type EGF-like homology <LEG2>
F:1776-2115/Domain: III <DOM3>
F:1776-1805/Domain: laminin-type EGF-like homology #status atypical <LE1>
F:1809-1856/Domain: laminin-type EGF-like homology <LE2>
F:1859-1914/Domain: laminin-type EGF-like homology <LE3>
F:1917-1967/Domain: laminin-type EGF-like homology <LE4>
F:1970-2014/Domain: laminin-type EGF-like homology <LE5>
F:2017-2061/Domain: laminin-type EGF-like homology <LE6>
F:2064-2109/Domain: laminin-type EGF-like homology <LE7>
F:2116-2697/Domain: I/II, heptad repeats <DOM2>
E:2698-3712/Domain: G <DOM3>
F:2698-2863/Domain: repeat G1 <RG1>
E:2864-3048/Domain: repeat G2 <RG2>
F:3049-3223/Domain: repeat G3 <RG3>
F:3079-3200/Domain: laminin G repeat homology <LG3>
E:3334-3528/Domain: repeat G4 <RG4>
E:3529-3712/Domain: repeat G5 <RG5>
F:1847-1850-1943-2024-2196-2215-2267-2301-2323-2482-2538-2569-2699-2720-2890-2933

	Query Match	13.8%	Score 77:	DB 2:	Length 3712;
	Best Local Similarity	23.9%	Pred. No. 31:		
	Matches	21,	Conservative	17,	Mismatches 20; Indels 30; Gaps 5.
Oy	23	CVCPCINRGVSTNAGDLATLTCQST-----QC-----PTGRMLD--DGVTID	63		
		: : : : : : : : : : : :			
Dd	492	CRACCNMGITINDCNAVTTGCECKLTNFGNCERCKHGVNYPPTCSYCDCCDNCTESE	551		
		: : : : : : : : : : : :			
Oy	64	VEDRSAAOCV-----KCKPNFY	80		
		: : : : : : : : : : : :			
Dd	552	ICNKSGGCGICREHGFGPRCDCLGIFY	579		
		: : : : : : : : : : : :			

RESULT 9.
A33502
major surface-labeled trophozoite antigen precursor - Giardia lamblia
C:Species: Giardia lamblia
C:Date: 09-Nov-1990 #sequence_revision 09-Nov-1990 #text_change 09-Sep-1997
C:Accession: A33502
R:Gillin, F.D.; Hagblom, P.; Harwood, J.; Alely, S.B.; Reiner, D.S.; McCaffery, M.; Se
Proc. Natl. Acad. Sci. U.S.A. 87, 4463-4467, 1990
A:Title: Isolation and expression of the gene for a major surface protein of Giardia
A:Reference number: A33502; MUID:90280395
A:Accession: A33502
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-713 <GILL>
A:Cross-references: GB:M33641; NID:g159131; PID:g159132
A:Keywords: surface antigen; transmembrane protein

	Query Match	13.5%	Score 75.5;	DB 2,	Length 713;
	Best Local Similarity	27.5%;	Pred. No 9.7;		
	Matches 28; Conservative	10;	Mismatches 35;	Indels 29;	Gaps 6.
QY	. 3 ACGENGAGNPPRANNAANGT-CVPCQINRRGSGVTNACDGLATLATQC-----	47			
Db	135 ACGDTTGVTIAGAAGTTGTGIADCAECSPAPDATAGAEAKRVAI-CKRGSKYLKDNVCVD	193			
QY	48 STCCPTGT-----ALDDGVTDVFEDRSAAQCYCKCFENFYNGS	84			
Db	194 KAQCNSTGNTNFAYVD-----SENGKRKCASCSDNL--NGG	227			

RESULT	10
B44018	
laminin B2t chain - human	

```

C:Species: Homo sapiens (man)
C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C:Accession: BA4018
R:Kallunki, P.; Sainio, K.; Eddy, R.; Byers, M.; Kallunki, T.; Sariola, H.; Beck, K.; H
J. Cell Biol. 119, 679-693, 1992
A>Title: A truncated laminin chain homologous to the B2 chain: structure, spatial expres
A:Reference number: A44018; MUID:93016279
A:Accession: BA4018
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1111 <KLU>
A:Cross-references: GB:215009; GB:S47253; NID:g34231; PIDN:CAA78729.1; PID:g34232
A:Experimental source: fibrosarcoma HT1080 cells
A>Note: sequence extracted from NCBI backbone (NCBIF:116637)
C:Superfamily: laminin-type EGF-like homology
F:84-128/Domain: laminin-type EGF-like homology <LEG1>
F:517-570/Domain: laminin-type EGF-like homology <LEG>

Query Match 13.4% Score 75; DB 2; Length 1111;
Best Local Similarity 26.7%; Pred. No. 16;
Matches 35; Conservative 11; Mismatches 25; Indels 60; Gaps 9;

QY 20 RGICVPCQINRGSGS-----VFNAG--DLATLAT 45
|:|:| | |
Db 78 RDRCLPCNNSNGSLSRCDNNGRCRKPGVYGACDRCLEPFHMLTDAAGCTQDRLDLS 137
|:|:| | |
QY 46 QCSSTCPTGTALDDCVTVDFDRSAQCYCKRP-----NFFYN--GSGPGGAPG 92
|:|:| | |
Db 138 KCDCC-PAGIA---GPCD-----AGRCV-CRPATVGERCDRCRSGYNYLDGNPEG--C 184
|:|:| | |
QY 93 VQVFAAGAAA 103
|:|:| | |
Db 185 TQCFYGHSSAS 195

RESULT 11
A44018
laminin B2 chain precursor, long form - human
M:Alternate names: cell-adhesive scatter factor ladsin; kalinin 105K chain; nicein 100K
C:Species: Homo sapiens (man)
C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C:Accession: A44018; S41097; A49401
R:Kallunki, P.; Sainio, K.; Eddy, R.; Byers, M.; Kallunki, T.; Sariola, H.; Beck, K.; H
J. Cell Biol. 119, 679-693, 1992
A>Title: A truncated laminin chain homologous to the B2 chain: structure, spatial expres
A:Reference number: A44018; MUID:93016279
A:Accession: A44018
A:Molecule type: mRNA
A:Residues: 1-1193 <KAL>
A:Cross-references: GB:215008; NID:g34229; PIDN:CAA78728.1; PID:g34230
A:Experimental source: fibrosarcoma HT1080 cells
A>Note: sequence extracted from NCBI backbone (NCBIF:116635)
R:Vailliy, J.; Verrando, P.; Champilaud, M.F.; Gerecke, D.; Wegman, D.W.; Baudoin, C.; A
Eur. J. Biochem. 219, 209-218, 1994
A>Title: The 100-kDa chain of nicein/kallinin is a laminin B2 chain variant.
A:Reference number: S41097; MUID:94139694
A:Accession: S41097
A:Molecule type: mRNA
A:Residues: 1-11, 'L', '13-472, 'M', '474-520, 'N', '522-856, 'R', '858-1193 <VAT>
A:Cross-references: EMBL:X73902; NID:g452754; PIDN:CAA52108.1; PID:g452755
A:Experimental source: keratinocytes
R:Miyaizaki, K.; Kikkawa, Y.; Nakamura, A.; Yasumitsu, H.; Umeda, M.
Proc. Natl. Acad. Sci. U.S.A. 90, 11767-11771, 1993
A>Title: A large cell-adhesive scatter factor secreted by human gastric carcinoma cells
A:Reference number: A49401; MUID:94089713
A:Accession: A49401
A:Molecule type: protein
A:Residues: 435-441, 'X', '443-444, 'X', '446-449 <MIY>
A:Gene: GDB:IAMC2; EBR2; EBR2A; L; LAMB2T
A:Cross-references: GDB:136225; OMIM:150292
A:Map position: 1q31-1q31

```

```

C:Superfamily:laminin-type EGF-like homology
C:Keywords: alternative splicing; basement membrane; extracellular matrix; glycoprotein
F:1-21/Domain: signal sequence #status predicted <IG>
F:22-193/Product: laminin B2t chain #status predicted <MAT>
F:84-128/Domain: laminin-type EGF-like homology <LEG1>
F:517-570/Domain: laminin-type EGF-like homology <LEG>

Query Match      13.4%   Score 75: DB 2: Length 1193:
Best Local Similarity 26.7%: Pred. No. 17:
Matches 35: Conservative 11: Mismatches 25: Indels 60: Gaps 9:

OY    20  RGICPCQINWVS-----VNAG---DLATLAT 45
       |::|::|||
Db     78  RDRCLPCCNCKSGISASACDNSGRCSCKPGVYGARCDRCPLGFMHMLTDAGCTGDORLDS 137
               :|::|::|
OY    46  QCSSTOCPTGTALDDGVYDFPDRSAQCVCAP-----NYNN--GSPQGEAPG 92
       :|::|::|
Db     138 KCDCD-PAGIA---GPCD-----AGRCV-CKPATYGERCDRCRGGYYVLDDGNPEG---C 184
               :|::|::|
OY    93  VQVFPAAGAAA 103
       |||::|:|
Db     185 TQCFGYGHSS 195

RESULT 12
MMSA
laminin alpha-1 chain precursor - mouse
N:Alternate names: laminin chain A1
C:Species: Mus musculus (house mouse)
C>Date: 30-Jun-1991 #sequence revision 30-Jun-1991 #text change 10-Dec-1999
C:Accession: A31771; A30449; S00624; A30450; S08895; S02678; S01790; A30451; S14670
J:Sasaki, M.; Kleinman, H.K.; Huber, H.; Deutzmann, R.; Yamada, Y.
J:BIOL. Chem. 265, 16536-16544, 1988
A>Title: Laminin, a multidomain protein. The A chain has a unique globular domain and
A:Reference number: A31771; MUID:89034134
A:Accession: A31771
A:Molecule type: mRNA
A:Residues: 1-3084 <SAS>
A:Cross-references: EMBL:J04064; NID:g309419; PIDN:AA39410.1; PID:g309420
A:Accession: A30449
A:Molecule type: protein
A:Residues: 183-195;570-571,'A',573-586;596-612,'X',614-617,'EMK';630-646;1217-1222,'
2486;2624-2639;2818-2843;3009-3033,'V',3035 <SRZ>
J:Hartl, L.; Oberbauer, I.; Deutzmann, R.
Eur. J. Biochem. 173, 629-635, 1988
A>Title: The N terminus of laminin A chain is homologous to the B chains.
A:Reference number: S00624; MUID:88225080
A:Accession: S00624
A:Molecule type: mRNA
A:Residues: 1-208,'T',210-334 <HAN>
A:Cross-references: EMBL:X07737; NID:g52857; PIDN:CA30561.1; PID:g52858
A:Accession: A30450
A:Molecule type: protein
A:Residues: 311-335,'N',337-339;630-642,'D',644;692-734;737-748,'X',750-760,'G',762-7
3-139;1449-1459 <HA>
A>Note: the sequence from Fig. 7 is inconsistent with that from Fig. 5 in having 209-
Eur. J. Biochem. 178, 71-80, 1988
R:Mann, K.; Deutzmann, R.; Timpl, R.
A>Title: Characterization of proteolytic fragments of the laminin-nidogen complex and
A:Reference number: S08895; MUID:89078415
A:Accession: S08895
A:Molecule type: protein
A:Residues: 153-169 <MAN>
R:Fujikawa, S.; Shinkai, H.; Deutzmann, R.; Paulsson, M.; Timpl, R.
Biochem. J. 252, 453-461, 1988
A>Title: Structure and distribution of N-linked oligosaccharide chains on various dom
A:Reference number: S02678; MUID:88326259
A:Accession: S02678
A:Molecule type: protein
A:Residues: 630-642,'D',644;2690-2704 <FUU>
Eur. J. Biochem. 177, 35-45, 1988
J:Deutzmann, R.; Huber, J.; Schmetz, K.A.; Oberbauer, I.; Hartl, L.

```

A:Title: Structural study of long arm fragments of laminin. Evidence for repetitive C-ter
 A:Reference number: S01790; MUID:89030693
 A:Accession: S01790
 A:Molecule type: mRNA
 A:Residues: 2538-3084 <DEU>
 A:Cross-references: EMBL:X13459; NID:955499; PIDN:CAA31807.1; PID:g818014
 A:Accession: A30451
 A:Molecule type: protein
 A:Residues: 1911-1929;1997-2006;2033-2045,'X',2047-2054,'X',2056-2066,'X',2068-2105;2120-2122;2124-2125;2127-2128;2130-2131;2133-2134;2136-2137;2139-2140;2142-2143;2145-2146;2148-2149;2151-2152;2154-2155;2157-2158;2160-2161;2163-2164;2166-2167;2169-2170;2172-2173;2175-2176;2178-2179;2181-2182;2184-2185;2187-2188;2190-2191;2193-2194;2196-2197;2199-2200;2202-2203;2205-2206;2208-2209;2211-2212;2214-2215;2217-2218;2220-2221;2223-2224;2226-2227;2229-2230;2232-2233;2235-2236;2238-2239;2241-2242;2244-2245;2247-2248;2250-2251;2253-2254;2256-2257;2259-2260;2262-2263;2265-2266;2268-2269;2271-2272;2274-2275;2277-2278;2280-2281;2283-2284;2286-2287;2289-2290;2292-2293;2295-2296;2298-2299;2301-2302;2304-2305;2307-2308;2310-2311;2313-2314;2316-2317;2319-2320;2322-2323;2325-2326;2328-2329;2331-2332;2334-2335;2337-2338;2340-2341;2343-2344;2346-2347;2349-2350;2352-2353;2355-2356;2358-2359;2361-2362;2364-2365;2367-2368;2370-2371;2373-2374;2376-2377;2379-2380;2382-2383;2385-2386;2388-2389;2391-2392;2394-2395;2397-2398;2400-2401;2403-2404;2406-2407;2409-2410;2412-2413;2415-2416;2418-2419;2421-2422;2424-2425;2427-2428;2430-2431;2433-2434;2436-2437;2439-2440;2442-2443;2445-2446;2448-2449;2451-2452;2454-2455;2457-2458;2460-2461;2463-2464;2466-2467;2469-2470;2472-2473;2475-2476;2478-2479;2481-2482;2484-2485;2487-2488;2490-2491;2493-2494;2496-2497;2499-2500;2502-2503;2505-2506;2508-2509;2511-2512;2514-2515;2517-2518;2520-2521;2523-2524;2526-2527;2529-2530;2532-2533;2535-2536;2538-2539;2541-2542;2544-2545;2547-2548;2550-2551;2553-2554;2556-2557;2559-2560;2562-2563;2565-2566;2568-2569;2571-2572;2574-2575;2577-2578;2580-2581;2583-2584;2586-2587;2589-2590;2592-2593;2595-2596;2598-2599;2601-2602;2604-2605;2607-2608;2610-2611;2613-2614;2616-2617;2619-2620;2622-2623;2625-2626;2628-2629;2631-2632;2634-2635;2637-2638;2640-2641;2643-2644;2646-2647;2649-2650;2652-2653;2655-2656;2658-2659;2661-2662;2664-2665;2667-2668;2670-2671;2673-2674;2676-2677;2679-2680;2682-2683;2685-2686;2688-2689;2691-2692;2694-2695;2697-2698;2700-2701;2703-2704;2706-2707;2709-2710;2712-2713;2715-2716;2718-2719;2721-2722;2724-2725;2727-2728;2730-2731;2733-2734;2736-2737;2739-2740;2742-2743;2745-2746;2748-2749;2751-2752;2754-2755;2757-2758;2760-2761;2763-2764;2766-2767;2769-2770;2772-2773;2775-2776;2778-2779;2781-2782;2783-2784;2786-2787;2789-2790;2792-2793;2795-2796;2798-2799;2801-2802;2804-2805;2807-2808;2810-2811;2813-2814;2816-2817;2819-2820;2822-2823;2825-2826;2828-2829;2831-2832;2834-2835;2837-2838;2840-2841;2843-2844;2846-2847;2849-2850;2852-2853;2854-2855;2857-2858;2860-2861;2863-2864;2866-2867;2869-2870;2872-2873;2875-2876;2878-2879;2881-2882;2883-2884;2886-2887;2889-2890;2892-2893;2895-2896;2898-2899;2901-2902;2904-2905;2907-2908;2910-2911;2913-2914;2916-2917;2919-2920;2922-2923;2925-2926;2928-2929;2931-2932;2934-2935;2937-2938;2940-2941;2943-2944;2946-2947;2949-2950;2952-2953;2955-2956;2958-2959;2961-2962;2964-2965;2967-2968;2970-2971;2973-2974;2976-2977;2979-2980;2982-2983;2985-2986;2988-2989;2991-2992;2994-2995;2997-2998;3000-3001;3003-3004;3006-3007;3009-3010;3012-3013;3015-3016;3018-3019;3021-3022;3024-3025;3027-3028;3030-3031;3033-3034;3036-3037;3039-3040;3042-3043;3045-3046;3048-3049;3051-3052;3054-3055;3057-3058;3060-3061;3063-3064;3066-3067;3069-3070;3072-3073;3075-3076;3078-3079;3081-3082;3084-3085;3087-3088;3090-3091;3093-3094;3096-3097;3099-3100;3102-3103;3105-3106;3108-3109;3111-3112;3114-3115;3117-3118;3120-3121;3123-3124;3126-3127;3129-3130;3132-3133;3135-3136;3138-3139;3141-3142;3144-3145;3147-3148;3150-3151;3153-3154;3156-3157;3159-3160;3162-3163;3165-3166;3168-3169;3171-3172;3174-3175;3177-3178;3180-3181;3183-3184;3186-3187;3189-3190;3192-3193;3195-3196;3198-3199;3201-3202;3204-3205;3207-3208;3210-3211;3213-3214;3216-3217;3219-3220;3222-3223;3225-3226;3228-3229;3231-3232;3234-3235;3237-3238;3240-3241;3243-3244;3246-3247;3249-3250;3252-3253;3255-3256;3258-3259;3261-3262;3264-3265;3267-3268;3270-3271;3273-3274;3276-3277;3279-3280;3282-3283;3285-3286;3288-3289;3291-3292;3294-3295;3297-3298;3300-3301;3303-3304;3306-3307;3309-3310;3312-3313;3315-3316;3318-3319;3321-3322;3324-3325;3327-3328;3330-3331;3333-3334;3336-3337;3339-3340;3342-3343;3345-3346;3348-3349;3351-3352;3354-3355;3357-3358;3360-3361;3363-3364;3366-3367;3369-3370;3372-3373;3375-3376;3378-3379;3381-3382;3384-3385;3387-3388;3390-3391;3393-3394;3396-3397;3399-3400;3402-3403;3405-3406;3408-3409;3411-3412;3414-3415;3417-3418;3420-3421;3423-3424;3426-3427;3429-3430;3432-3433;3435-3436;3438-3439;3441-3442;3444-3445;3447-3448;3450-3451;3453-3454;3456-3457;3459-3460;3462-3463;3465-3466;3468-3469;3471-3472;3474-3475;3477-3478;3480-3481;3482-3483;3485-3486;3488-3489;3491-3492;3494-3495;3497-3498;3500-3501;3503-3504;3506-3507;3509-3510;3512-3513;3515-3516;3518-3519;3521-3522;3524-3525;3527-3528;3530-3531;3533-3534;3536-3537;3539-3540;3542-3543;3545-3546;3548-3549;3551-3552;3554-3555;3557-3558;3560-3561;3562-3563;3565-3566;3568-3569;3571-3572;3574-3575;3577-3578;3580-3581;3582-3583;3585-3586;3588-3589;3591-3592;3594-3595;3597-3598;3600-3601;3603-3604;3606-3607;3609-3610;3612-3613;3615-3616;3618-3619;3621-3622;3624-3625;3627-3628;3630-3631;3633-3634;3636-3637;3639-3640;3642-3643;3645-3646;3648-3649;3651-3652;3654-3655;3657-3658;3660-3661;3663-3664;3666-3667;3669-3670;3672-3673;3675-3676;3678-3679;3681-3682;3684-3685;3687-3688;3690-3691;3693-3694;3696-3697;3699-3700;3702-3703;3705-3706;3708-3709;3711-3712;3714-3715;3717-3718;3720-3721;3723-3724;3726-3727;3729-3730;3732-3733;3735-3736;3738-3739;3741-3742;3744-3745;3747-3748;3750-3751;3753-3754;3756-3757;3759-3760;3762-3763;3765-3766;3768-3769;3771-3772;3774-3775;3777-3778;3780-3781;3783-3784;3786-3787;3789-3790;3792-3793;3795-3796;3798-3799;3801-3802;3804-3805;3807-3808;3810-3811;3813-3814;3816-3817;3819-3820;3822-3823;3825-3826;3828-3829;3831-3832;3834-3835;3837-3838;3840-3841;3843-3844;3846-3847;3849-3850;3852-3853;3854-3855;3857-3858;3860-3861;3863-3864;3866-3867;3869-3870;3872-3873;3875-3876;3878-3879;3881-3882;3884-3885;3887-3888;3890-3891;3893-3894;3896-3897;3899-3900;3902-3903;3905-3906;3908-3909;3911-3912;3914-3915;3917-3918;3920-3921;3923-3924;3926-3927;3929-3930;3932-3933;3935-3936;3938-3939;3941-3942;3944-3945;3947-3948;3950-3951;3953-3954;3956-3957;3959-3960;3962-3963;3965-3966;3968-3969;3971-3972;3974-3975;3977-3978;3980-3981;3983-3984;3986-3987;3989-3990;3992-3993;3995-3996;3998-3999;4001-4002;4004-4005;4007-4008;4010-4011;4013-4014;4016-4017;4019-4020;4022-4023;4025-4026;4028-4029;4031-4032;4034-4035;4037-4038;4040-4041;4043-4044;4046-4047;4049-4050;4052-4053;4055-4056;4058-4059;4061-4062;4063-4064;4066-4067;4069-4070;4072-4073;4075-4076;4078-4079;4081-4082;4084-4085;4087-4088;4090-4091;4093-4094;4096-4097;4099-4100;4102-4103;4105-4106;4108-4109;4111-4112;4114-4115;4117-4118;4120-4121;4123-4124;4126-4127;4129-4130;4132-4133;4135-4136;4138-4139;4141-4142;4144-4145;4147-4148;4150-4151;4153-4154;4156-4157;4159-4160;4162-4163;4165-4166;4168-4169;4171-4172;4174-4175;4177-4178;4180-4181;4183-4184;4186-4187;4189-4190;4192-4193;4195-4196;4198-4199;4201-4202;4204-4205;4207-4208;4210-4211;4213-4214;4216-4217;4219-4220;4222-4223;4225-4226;4228-4229;4231-4232;4234-4235;4237-4238;4240-4241;4243-4244;4246-4247;4249-4250;4252-4253;4255-4256;4258-4259;4261-4262;4264-4265;4267-4268;4270-4271;4273-4274;4276-4277;4279-4280;4282-4283;4285-4286;4288-4289;4291-4292;4294-4295;4297-4298;4300-4301;4303-4304;4306-4307;4309-4310;4312-4313;4315-4316;4318-4319;4321-4322;4324-4325;4327-4328;4330-4331;4333-4334;4336-4337;4339-4340;4342-4343;4345-4346;4348-4349;4351-4352;4354-4355;4357-4358;4360-4361;4363-4364;4366-4367;4369-4370;4372-4373;4375-4376;4378-4379;4381-4382;4384-4385;4387-4388;4390-4391;4393-4394;4396-4397;4399-4400;4402-4403;4405-4406;4408-4409;4411-4412;4414-4415;4417-4418;4420-4421;4423-4424;4426-4427;4429-4430;4432-4433;4435-4436;4438-4439;4441-4442;4444-4445;4447-4448;4450-4451;4453-4454;4456-4457;4459-4460;4462-4463;4465-4466;4468-4469;4471-4472;4474-4475;4477-4478;4480-4481;4482-4483;4485-4486;4488-4489;4491-4492;4494-4495;4497-4498;4500-4501;4502-4503;4505-4506;4508-4509;4510-4511;4513-4514;4516-4517;4519-4520;4522-4523;4525-4526;4528-4529;4531-4532;4534-4535;4537-4538;4540-4541;4543-4544;4546-4547;4549-4550;4552-4553;4554-4555;4557-4558;4560-4561;4562-4563;4565-4566;4568-4569;4571-4572;4574-4575;4577-4578;4580-4581;4582-4583;4585-4586;4588-4589;4591-4592;4594-4595;4597-4598;4600-4601;4603-4604;4606-4607;4609-4610;4612-4613;4615-4616;4618-4619;4621-4622;4624-4625;4627-4628;4630-4631;4632-4633;4635-4636;4638-4639;4641-4642;4644-4645;4647-4648;4650-4651;4653-4654;4656-4657;4658-4659;4661-4662;4664-4665;4667-4668;4670-4671;4673-4674;4676-4677;4679-4680;4682-4683;4685-4686;4688-4689;4691-4692;4694-4695;4697-4698;4700-4701;4702-4703;4705-4706;4708-4709;4711-4712;4714-4715;4717-4718;4720-4721;4723-4724;4726-4727;4729-4730;4732-4733;4735-4736;4738-4739;4741-4742;4744-4745;4747-4748;4750-4751;4753-4754;4756-4757;4759-4760;4762-4763;4765-4766;4768-4769;4771-4772;4774-4775;4777-4778;4780-4781;4783-4784;4786-4787;4789-4790;4792-4793;4795-4796;4798-4799;4801-4802;4804-4805;4807-4808;4810-4811;4813-4814;4816-4817;4819-4820;4822-4823;4825-4826;4828-4829;4831-4832;4834-4835;4837-4838;4840-4841;4843-4844;4846-4847;4849-4850;4852-4853;4854-4855;4857-4858;4860-4861;4863-4864;4866-4867;4869-4870;4872-4873;4875-4876;4878-4879;4881-4882;4884-4885;4887-4888;4890-4891;4893-4894;4896-4897;4899-4900;4902-4903;4905-4906;4908-4909;4910-4911;4913-4914;4916-4917;4919-4920;4922-4923;4925-4926;4928-4929;4931-4932;4934-4935;4937-4938;4940-4941;4943-4944;4946-4947;4949-4950;4952-4953;4954-4955;4957-4958;4960-4961;4963-4964;4966-4967;4969-4970;4972-4973;4975-4976;4978-4979;4981-4982;4984-4985;4987-4988;4990-4991;4993-4994;4996-4997;4999-5000;5002-5003;5005-5006;5008-5009;5010-5011;5013-5014;5016-5017;5019-5020;5022-5023;5025-5026;5028-5029;5031-5032;5034-5035;5037-5038;5040-5041;5043-5044;5046-5047;5049-5050;5052-5053;5054-5055;5057-5058;5060-5061;5063-5064;5066-5067;5069-5070;5072-5073;5075-5076;5078-5079;5081-5082;5084-5085;5087-5088;5090-5091;5093-5094;5096-5097;5099-5100;5102-5103;5105-5106;5108-5109;5111-5112;5114-5115;5117-5118;5120-5121;5123-5124;5126-5127;5129-5130;5132-5133;5135-5136;5138-5139;5141-5142;5144-5145;5147-5148;5150-5151;5153-5154;5156-5157;5159-5160;5162-5163;5165-5166;5168-5169;5171-5172;5174-5175;5177-5178;5180-5181;5183-5184;5186-5187;5189-5190;5192-5193;5195-5196;5198-5199;5201-5202;5204-5205;5207-5208;5210-5211;5213-5214;5216-5217;5219-5220;5222-5223;5225-5226;5228-5229;5231-5232;5234-5235;5237-5238;5240-5241;5243-5244;5246-5247;5249-5250;5252-5253;5255-5256;5258-5259;5261-5262;5264-5265;5267-5268;5270-5271;5273-5274;5276-5277;5279-5280;5282-5283;5285-5286;5288-5289;5291-5292;5294-5295;5297-5298;5300-5301;5303-5304;5306-5307;5309-5310;5312-5313;5315-5316;5318-5319;5321-5322;5324-5325;5327-5328;5330-5331;5333-5334;5336-5337;5339-5340;5342-5343;5345-5346;5348-5349;5351-5352;5354-5355;5357-5358;5360-5361;5363-5364;5366-5367;5369-5370;5372-5373;5375-5376;5378-5379;5381-5382;5384-5385;5387-5388;5390-5391;5393-5394;5396-5397;5399-5400;5402-5403;5405-5406;5408-5409;5411-5412;5414-5415;5417-5418;5420-5421;5423-5424;5426-5427;5429-5430;5432-5433;5435-5436;5438-5439;5441-5442;5444-5445;5447-5448;5450-5451;5453-5454;5456-5457;5459-5460;5462-5463;5465-5466;5468-5469;5471-5472;5474-5475;5477-5478;5480-5481;5482-5483;5485-5486;5488-5489;5491-5492;5494-5495;5497-5498;5500-5501;5502-5503;5505-5506;5508-5509;5510-5511;5513-5514;5516-5517;5519-5520;5522-5523;5525-5526;5528-5529;5531-5532;5534-5535;5537-5538;5540-5541;5543-5544;5546-5547;5549-5550;5552-5553;5554-5555;5557-5558;5560-5561;5562-5563;5565-5566;5568-5569;5571-5572;5574-5575;5577-5578;5580-5581;5582-5583;5585-5586;5588-5589;5591-5592;5594-5595;5597-5598;5600-5601;5603-5604;5606-5607;5609-5610;5612-5613;5615-5616;5618-5619;5621-5622;5624-5625;5627-5628;5630-5631;5632-5633;5635-5636;5638-5639;5641-5642;5644-5645;5647-5648;5650-5651;5653-5654;5656-5657;5658-5659;5661-5662;5664-5665;5667-5668;5670-5671;5673-5674;5676-5677;5679-5680;5682

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:48:00 ; Search time 116.78 Seconds
(without alignments)
105.385 Million cell updates/sec

Title: US-09-196-161d-1
Perfect score: 558
Sequence: 1 GAAQGEANGOPFANNAAR.....POGEAPGVQVFAAGAAAGV 105

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: SPREMBL_15:*
 - 2: sp_archaea:*
 - 3: sp_bacteria:*
 - 4: sp_fungi:*
 - 5: sp_human:*
 - 6: sp_invertebrate:*
 - 7: sp_mammal:*
 - 8: sp_mmc:*
 - 9: sp_organelle:*
 - 10: sp_phage:*
 - 11: sp_plant:*
 - 12: sp_podent:*
 - 13: sp_virus:*
 - 14: sp_vertebrate:*
 - 15: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	558	100.0	394	5	Q27208
2	558	100.0	442	5	Q9XZ62
3	84.5	15.1	1737	4	Q75097
4	80.5	14.4	560	5	Q9U013
5	79.5	14.2	874	11	Q9QYPO
6	79.5	14.2	1019	5	Q9NA40
7	79	14.2	600	13	Q9PVK7
8	78	13.8	597	5	Q07317
9	77	13.8	424	5	Q9XZ41
10	77	13.8	634	5	Q9VRM0
11	77	13.8	1372	5	P91526
12	77	13.8	1546	4	Q9NS27
13	77	13.8	1551	4	Q75445
14	76.5	13.7	400	13	Q9P447
15	76	13.6	1725	11	P70570
16	75.5	13.5	1696	5	Q9NJ15
17	75	13.4	620	5	Q9XZX8
18	74.5	13.4	620	13	O42138
19	74	13.3	594	5	Q24970

20	72.5	13.0	398	1	Q00388	Q00388 methanococc
21	72.5	13.0	424	5	Q76744	Q76744 necator ame
22	72.5	13.0	2225	5	Q45881	Q45881 caenorhabdi
23	72	12.9	167	5	Q24991	Q24991 giardia lam
24	72	12.9	468	4	Q00220	Q00220 homo sapien
25	71.5	12.8	498	11	Q60846	Q60846 mus musculu
26	70.5	12.6	289	5	P91237	P91237 caenorhabdi
27	70.5	12.6	425	5	Q71153	Q71153 ancylostoma
28	70.5	12.6	598	5	Q9N848	Q9N848 leishmania
29	70.5	12.6	709	5	Q9XZX9	Q9XZX9 leishmania
30	70	12.5	1792	13	O57484	O57484 gallus gall
31	69.5	12.5	401	5	Q9NH87	Q9NH87 giardia lam
32	69.5	12.5	2106	2	Q9XC47	Q9XC47 rickettsia
33	69.5	12.5	2972	5	P90891	P90891 caenorhabdi
34	69	12.4	1086	10	Q9LIT3	Q9LIT3 zea mays (m
35	68.5	12.3	331	6	Q9TSN4	Q9TSN4 macaca fasc
36	68.5	12.3	371	5	Q24990	Q24990 giardia lam
37	68.5	12.3	425	5	Q19348	Q19348 caenorhabdi
38	68.5	12.3	484	2	O51683	O51683 paracoccus
39	68.5	12.3	580	4	Q9P2M2	Q9P2M2 homo sapien
40	68.5	12.3	956	4	Q9Y4H1	Q9Y4H1 homo sapien
41	68.5	12.3	956	4	Q9UEJ7	Q9UEJ7 homo sapien
42	68.5	12.3	962	4	Q9Y4H0	Q9Y4H0 homo sapien
43	68.5	12.3	962	4	Q9UEJ8	Q9UEJ8 homo sapien
44	68.5	12.3	969	4	Q9UEJ7	Q9UEJ7 homo sapien
45	68.5	12.3	975	4	Q9Y4G9	Q9Y4G9 homo sapien

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	394 AA.
Q27208	Q27208	Q27208		
AC	Q27208	Q27208		
DT	01-NOV-1996 (TREMBLER, 01, Created)			
DT	01-NOV-1998 (TREMBLER, 08, Last sequence update)			
DT	01-JUN-2000 (TREMBLER, 14, Last annotation update)			
DE	IMMOBILIZATION ANTIGEN PRECURSOR (FRAGMENT).			
OS	Ichthyophthirius multifiliis.			
OC	Ichthyophthirius multifiliis.			
CC	Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;			
CC	Opisthokonta; Ichthyophthirius.			
OX	NCBI_TaxID:5932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=GEORGIA;			
RX	MEDLINE=92335298; PubMed=1631132;			
RA	Clark T.G., McGraw R.A., Dickerson H.W.;			
RT	"Purification and partial characterization of immobilization antigens			
RT	ciliate Ichthyophthirius multifiliis.";			
RT	Proc. Natl. Acad. Sci. U.S.A. 89:6363-6367(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=GEORGIA;			
RX	MEDLINE=93020590; PubMed=1383510;			
RA	Liu T.L., Dickerson H.W.;			
RT	"Purification and partial characterization of immobilization antigens			
RT	from Ichthyophthirius multifiliis.";			
J	Protocols. 39:457-463(1992).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=GEORGIA;			
RA	Clark T.;			
RT	Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=GEORGIA;			
RA	Clark T.;			
RT	Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: M92907; AAC36158.1; -.			
KW	Signal.			
FT	NON-TER	1	1	POTENTIAL:
FT	SIGNAL	<1	1	

	Best Local Similarity	26.8%;	Pred. No.	0.78;						
	Matches	26;	Conservative	8;	Mismatches	38;	Indels	25;	Gaps	4;
OY	14 AANNARG---	ICVPCQI----	NRVGS-----	TNAGDLATLATQCSTGCTGTALDGG	60					
Dd	224 ACNTAANGIDKCAECISLTPASRGAILITCTKSTNSLSPLKDACLTCSPAGTYETGS	283								
OY	61 VTDVF-----	-DRSAACVCKCKPNFYNGS	85							
Dd	284 PNKVCTPCHTSCAGCKNDNTASTACTACPGSVLSYGS	320								
RESULT	5									
O9OYP0		PRELIMINARY;	PRT;	874 AA.						
ID	O9OYP0									
AC	O9OYP0;									
DT	01-MAY-2000 (TREMBLrel. 13, Created)									
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)									
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)									
DE	MEGF8 (FRAGMENT).									
GN	MEGF8.									
OS	Rattus norvegicus (Rat).									
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;									
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.									
OX	NCBI_TaxId=10116;									
RN	[1]									
RP	SEQUENCE FROM N.A.									
RA	STRAIN=SPRACOE-Dawley; TISSUE=BRAIN;									
RC	Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;									
RT	"Identification of high-molecular-weight proteins with multiple Egf-									
RL	like motifs by motif-trap screening.";									
Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.										
EMBL; AB011534; BAA8689.1; -										
HSSP; P35555; IEMN.										
DR INTERPRO; IPRO000152; -										
DR INTERPRO; IPR000561; -										
DR INTERPRO; IPR001881; -										
DR INTERPRO; IPR002049; -										
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.										
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.										
DR PROSITE; PS01186; EGF_2; 2.										
DR PROSITE; PS01248; LAMININ_TYPE_EGF; UNKNOWN_2.										
FT NON_TER	1									
SQ SEQUENCE	874 AA;	93764 MW;	CB63BDF183D870AB	CRC64;						
Query Match		14.2%;	Score	79.5;	DB	11;	Length	874;		
Best Local Similarity		32.4%;	Pred. No.	1.6;						
Matches	24;	Conservative	4;	Mismatches	29;	Indels	17;	Gaps	3;	
OY	21 GICVPCQINRVGSVTNAGDLATLATQCSTGCP--	-TGTLADGVIIVDFDRSAACGVK	74							
Dd	404 GKCRFKCQCN-----	-GHADTCNQDDQTGCPCQNNTETVGCGSSPSDRNRCKYKCGAK	455							
OY	75 CKPNFYNGSGSPQG	88								
	: :									
Dd	456 CRESEH---	-GSLPG	466							
RESULT	6									
O9NA40		PRELIMINARY;	PRT;	1019 AA.						
ID	O9NA40									
AC	O9NA40;									
DT	01-OCT-2000 (TREMBLrel. 15, Created)									
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)									
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)									
DE	V73F8A.5 PROTEIN.									
GN	V73F8A.5.									
OS	Caenorhabditis elegans.									
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;									
CC	Rhabditidae; Pelodermidae; Caenorhabditis.									
OX	NCBI_TaxId=6239;									
NN	[1]									

```

RP      SEQUENCE FROM N.A.
RA      Matthews L.;
RL      Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN      12]
RR      none;
RW      SEQUENCE FROM N.A.
RT      "Genome sequence of the nematode C.elegans: A platform for
        investigating biology."
RL      Science 282:2012-2018(1998).
DR      EMBL; AL162862; CAB70224.1; -.
SQ      SEQUENCE 1019 AA; 112378 MW;  6EB4094722B707E7 CRC64;

Query Match                               14.2%; Score 79.5; DB 5; Length 1019;
Best Local Similarity                      29.9%; Pred. No. 1.9;
Matches 26; Conservative 12; Mismatches 26; Indels 23; Gaps 5;

OY      10 NOPFANNARICVPCQINRYGVTNMGDLATLA-TCCSTQCPTGTALDDGVTFDRS 68
Db       242 NNGYASNO-----PIHNDRI-----DVLGLAFTRCTACPGTSSPGG----- 280
          |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY      69 AAOVCCKKPNFYNGSPQ-GEAPGVQ 94
Db       281 SAECIPCSPGSSSKSGCGCRCPESQ 307
          ::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 7
O9PVK7 PRELIMINARY; PRT; 600 AA.
ID O9PVK7
AC O9PVK7;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE COHRIN PRECURSOR.
OS Naja naja (Indian cobra).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodonta; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Elapinae; Naja.
OX NCBI_TaxID=35670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KAOUTHIA; TISSUE-VENOM GLAND;
RC Bandai B., Bredehorst R., Vogel C.-W.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF063190; AAF00693.1; -.
DR HSSP; P17494; 1KST.
DR INTERPRO; IPRO00130; -.
DR INTERPRO; IPRO01590; -.
DR INTERPRO; IPRO01762; -.
DR INTERPRO; IPRO02870; -.
DR PFAM; PF00200; disintegrin_1.
DR PFAM; PF01421; Repolyasin_1.
DR PFAM; PF01562; pep_M12B_propep_1.
DR PRINTS; PR00289; DISINTEGRIN.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW signal.
DR SIGNAL.
FT CHAIN 180 600 POTENTIAL.
FT CHAIN 180 600 COBRIN.
SO SEQUENCE 600 AA; 67661 MW; C7C0D45EBC694290 CRC64;

Query Match                               14.2%; Score 79; DB 13; Length 600;
Best Local Similarity                     31.8%; Pred. No. 1.2;
Matches 27; Conservative 7; Mismatches 27; Indels 24; Gaps 5;

OY      23 CVPCQINRVGSVTNAG---DLATLATQCSFGCPTGTALDGDGYTVDFRSAAGCVCKRPN 78
Db       442 CEKKFKAGAGAECHRAKDCCDLPELTQTQSACEP-----TDVFGRNG---LPQNN 489
          |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY      79 FY-YNIGSP-----QGEAPGVQ 95
Db       490 GYCYNRGCPIMTNOCIALRGPGVKV 514
          |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

	Query Match	14.0%;	Score 78;	DB 5;	Length 597;	
	Best Local Similarity	23.0%;	Pred. NO. 1.6;			
	Matches 28;	Conservative 10;	Mismatches 36;	Indels 48;	Gaps 6;	
OY	9 GNOPPANNAAAG---ICVPCQINRVGSVTNAGD-----LAT 42	: : :				
Db	237 GKTHPTTDSAGNNKKVCVSCGCTTNNGGLENGECTSKESARAGTEITCTCCSSNNLSP 296	: : :				
OY	43 LATCGSTCPTGTALDDG-----VTDFRGAACVKCKPNF--YYNG 83	: : :				
Db	297 LGDACTTCDCPACTAYAVSGSDGSVCVKPCRHNTCAQCQTD--DRFTS-CTACYPGSYLLYEEN 353	: : :				
OY	84 GS 85					
Db	354 GA 355					
RESULT	9					
O9XZ41						
ID	O9XZ41	PRELIMINARY;	PRT:	424 AA.		
AC	O9XZ41;					
DT	01-NOV-1999 (TREMBLrel. 12, Created)					
DT	01-NOV-1999 (TREMBLrel. 12, last sequence update)					
DT	01-OCT-2000 (TREMBLrel. 15, last annotation update)					
DE	ANCYLOSTOMA-SECRETED PROTEIN 1 PRECURSOR.					
OS	Ancylostoma caninum (Dog hookworm).					
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Strongylida;					
NCBI_TaxID=29170;	Ancylostomatidae; Ancylostomatinae; Ancylostoma.					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RA	STRAIN-SHANGHAI;					
RT	Zhan B., Shan Q., Hawdon J.M.;					
RT	"variation between ASP-1 molecules from Ancylostoma caninum in China and the US."					

[illegible]

RESULT	10			
09VRMO	ID	09VRMO	PRELIMINARY;	PRT; 634 AA.
AC	09VRMO;			
DT	01-MAY-2000 (TREMBlrel. 13, Created)			
DT	01-MAY-2000 (TREMBlrel. 13, Last sequence update)			
DT	01-OCT-2000 (TREMBlrel. 15, Last annotation update)			
DE	LANA PROTEIN.			
GN	LANA.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxId=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BERKELEY;			
RX	MEDLINE=20196006; PubMed=10731132;			
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
RA	Georgie R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sutton G.G., Wortman J.R., Tandell M.D., Zhang Q., Chen L.X.,			
RA	Brandon R.C., Rogers Y.-H.C., Blazel R.G., Champe M., Pfeiffer B.D.,			
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,			
RA	Abbil J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,			
RA	Ballew R.F., Basu A., Baxendale J., Bayraktaroglu I., Beasley E.M.,			
RA	Beeson K.Y., Benos P.V., Berner B.P., Bhandari D., Bolshakov S.,			
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,			
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA	Cherry J.M., Cawley S., Dahlke S., Davenport L.B., Davies P.,			
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,			
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,			
RA	Foster C., Gabriellian A.E., Garg N.S., Gelbatt W.M., Glasser K.,			
RA	Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,			
RA	Hosain D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,			
RA	Jalili M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,			
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,			
RA	Lau X.P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,			
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,			
RA	Mekunlu G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,			
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,			
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,			
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,			
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,			

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtks R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003563; AAF50672.1; -
 DR HSSP: P02468; ITRF;
 DR FLIBASE: FBgn002526; LANA.
 DR INTERPRO: IPR000561; -
 DR INTERPRO: IPR001886; -
 DR INTERPRO: IPR002049; -
 DR PFAM: PF00053; laminin_EGF_7.
 DR PFAM: PF00055; laminin_Nterm; 1.
 DR PRINTS: PR00011; EGF_LAMININ.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_5.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; 6.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; 6.
 SQ SEQUENCE 634 AA: 70927 MW: 51FC7DB3D08E9449 CRC64;

Query Match 13.8%; Score 77; DB 5; Length 634;
 Best Local Similarity 23.9%; Pred. No. 2.1;
 Matches 21; Conservative 17; Mismatches 20; Indels 30; Gaps 5;

OY 23 CVPQINVGSTNAGDLATLATGCGT-----QC-----TGTALD---DGV-TD 63
 DB 492 CACCCNCGISGTTDNCVTGCKCLTNGGDCNCERCKHGFYNPYCGSYCDNOGTESE 551
 OY 64 VEDRSAQC-----KCKPNFY 80
 DB 552 ICNKSGCICREGRGRCDCQCLPGFY 579

RESULT 11

ID P91526 PRELIMINARY; PRT: 1372 AA.
 AC P91526;
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE SIMILARITY TO MULTIPLE EGF-LIKE DOMAINS.
 GN W02C12.1.
 OS *Caenorhabditis elegans*.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;
 OC Rhabditiidae; Pelodierinae; *Caenorhabditis*.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA MEDLIN=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Koopra A., Saunders D., Showkhen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*."
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Murray J., Woldmann P.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.

RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U80815; AAB37995.1; -
 DR HSSP: Q12784; ITRF;
 DR INTERPRO: IPR000152; -
 DR INTERPRO: IPR000561; -
 DR INTERPRO: IPR000742; -
 DR INTERPRO: IPR001438; -
 DR INTERPRO: IPR001881; -
 DR PFAM: PF00008; EGF_12.
 DR PRINTS: PR00010; EGF_BLOOD.
 DR PROSITE: PS00010; ASX_HYDROXYL; 5.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_11.
 DR PROSITE: PS01186; EGF_2; 8.
 DR PROSITE: PS01187; EGF_CA; 2.
 KW Glycoprotein; EGF-like domain.
 SQ SEQUENCE 1372 AA: 150561 MW: 0537FA79BA0576C5 CRC64;

Query Match 13.8%; Score 77; DB 5; Length 1372;
 Best Local Similarity 21.2%; Pred. No. 4.7;
 Matches 28; Conservative 14; Mismatches 50; Indels 40; Gaps 5;

OY 3 AAGEANGNQPEFANNAAGICVPCQINRVS-----VTNAGDLATLATGCGT 49
 DB 1036 SEDECRDCPCPOEQLSASGVCPQIGITRSGEKKCYACPGTITATMTREQCNT 1095
 OY 50 Q-----CPTGTALDDGVTVEDRSAQCCKKPNFYNGSSPOGEAPG 92
 DB 1096 PKCKPQFLVETKNCQCPGRTFON-----EEQESTCKLCAPD--HTTAP-GATAE 1145
 OY 93 VQVFAAGAAAG 104
 DB 1146 SQCESTNOCATG 1157

RESULT 12

ID O9NS27 PRELIMINARY; PRT: 1546 AA.
 AC O9NS27;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE USHER SYNDROME TYPE IIA PROTEIN.
 OS *Homo sapiens* (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Weston M.D., Eudy J.D., Fugita T., Yao S.-F., Usami S., Creemers C.,
 RA Greenburg J., Ramasar R., Martini A., Moller C., Smith R.J.,
 RA Sunegul J., Kimberling W.J.;
 RT "The Gene Structure and Identification of Novel Mutations in *Usher1n*,
 RT the Gene Responsible for *Usher* Syndrome Type IIA.";
 RL Am. J. Hum. Genet. 66:0-0(2000).
 DR EMBL: AF091889; AAF75819.1; -
 DR EMBL: AF091873; AAF75819.1; JOINED.
 DR EMBL: AF091874; AAF75819.1; JOINED.
 DR EMBL: AF091875; AAF75819.1; JOINED.
 DR EMBL: AF091876; AAF75819.1; JOINED.
 DR EMBL: AF091877; AAF75819.1; JOINED.
 DR EMBL: AF091878; AAF75819.1; JOINED.
 DR EMBL: AF091879; AAF75819.1; JOINED.
 DR EMBL: AF091880; AAF75819.1; JOINED.
 DR EMBL: AF091881; AAF75819.1; JOINED.
 DR EMBL: AF091882; AAF75819.1; JOINED.
 DR EMBL: AF091883; AAF75819.1; JOINED.
 DR EMBL: AF091884; AAF75819.1; JOINED.
 DR EMBL: AF091885; AAF75819.1; JOINED.

DR EMBL: AF091886; AAF75819.1; JOINED.
DR EMBL: AF091887; AAF75819.1; JOINED.
DR EMBL: AF091888; AAF75819.1; JOINED.
SQ SEQUENCE 1546 AA; 171001 MW; 64BD43596C1E17B CRC64;

Query Match 13.8%; Score 77; DB 4; Length 1546;
Best Local Similarity 20.8%; Pred. No. 5.4;
Matches 26; Conservative 17; Mismatches 34; Indels 48; Gaps 5;

QY 9 GNOFPAANNAARGICVPCQINRVGSVTNAGDLATLATQCSYQCP-----TGT----- 55
DB 832 GNFYLRQNNSE--LCLEPCNCKDTGTING-----SLLCNKSSTGQCPCKLGVTLGRNQCEPH 885
QY 56 -----ALDDGVTD-----VFDSAAQCVCCKRNFYNGGSP 86
DB 886 RYNLTIDNFQHCQMCEDSLGTLPGLTICDPISGQCLCVPNQGRNCQCGFYISPGNA 945
QY 87 QGEAP 91
DB 946 TGCLP 950

RESULT 13

075445 PRELIMINARY; PRT: 1551 AA.

ID 075445;
AC 075445;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE USHER SYNDROME TYPE IIA PROTEIN.
GN USH2A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98288356; PubMed=9624053;
RA Eddy J.D., Weston M.D., Yao S.F., Hoover D.M., Rehm H.L., Ahmad I.,
RA Mc-Edmonds M., Yan D., Cheng J.J., Belsel K.W., Ayuso C., Creemers C.,
RA Davenport S., Moller G., Talmadge C.B., Tamayo M., Swarcop A.,
RA Morton C.C., Kimberling W.J., Sumegi J.;
RT "Mutation of a gene encoding a protein with extracellular matrix
motifs in Usher syndrome type IIA";
RL Science 280:1753-1757(1998).
DR EMBL: AF055580; AAC23748.1; -;
DR HSSP: P02468; IKLO.
DR INTERPRO: IPR000561; -;
DR INTERPRO: IPR001777; -;
DR INTERPRO: IPR001886; -;
DR INTERPRO: IPR002049; -;
DR PFAM: PF00041; fn3; 4.
DR PFAM: PF00053; laminin_EGF; 10.
DR PFAM: PF00055; laminin_Nterm; 1.
DR PRINTS: PR00011; EGF_LAMININ.
DR PROSITE: PS00022; EGF_1; UNKNOWN_7.
DR PROSITE: PS01248; LAMININ_TYPE_EGF; 7.
DR PRODOM: PD002082; -; 1.
SQ SEQUENCE 1551 AA; 171590 MW; FFA765CD0ACAC6B CRC64;

Query Match 13.8%; Score 77; DB 4; Length 1551;
Best Local Similarity 20.8%; Pred. No. 5.4;
Matches 26; Conservative 17; Mismatches 34; Indels 48; Gaps 5;

QY 9 GNOFPAANNAARGICVPCQINRVGSVTNAGDLATLATQCSYQCP-----TGT----- 55
DB 837 GNFYLRQNNSE--LCLEPCNCKDTGTING-----SLLCNKSSTGQCPCKLGVTLGRNQCEPH 890
QY 56 -----ALDDGVTD-----VFDSAAQCVCCKRNFYNGGSP 86

DB 891 RYNLTIDNFQHCQMCEDSLGTLPGLTICDPISGQCLCVPNQGRNCQCGFYISPGNA 950
QY 87 QGEAP 91
DB 951 TGCLP 955

RESULT 14

09PT47 PRELIMINARY; PRT: 400 AA.

ID 09PT47;
AC 09PT47;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE METALLOPROTEINASE PRECURSOR (FRAGMENT).
OS Atractaspis engadensis (Israeli burrowing asp).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Atractaspididae; Atractaspis.
OX NCBI_TaxID=8600;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=VENOM GLAND;
RA Duncanel F.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF186370; AAF01042.1; -;
DR HSSP: P17494; IKST.
DR INTERPRO: IPR00130; -;
DR INTERPRO: IPR001590; -;
DR INTERPRO: IPR001762; -;
DR PFAM: PF00200; disintegrin; 2.
DR PFAM: PF01421; Reptolysin; 1.
DR PRINTS: PR00289; DISINTEGRIN.
DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
FT NON_TER 1 203 METALLOPROTEINASE.
FT CHAIN 1 204 DISINTEGRIN.
FT CHAIN 204 280
SQ SEQUENCE 400 AA; 44790 MW; B562FAFD2EB79468 CRC64;

Query Match 13.7%; Score 76.5; DB 13; Length 400;
Best Local Similarity 23.1%; Pred. No. 1.5;
Matches 25; Conservative 10; Mismatches 32; Indels 41; Gaps 4;

QY 10 GNOFPAANNAARGIC-----VPCQINRVGSVTNAG----- 38
DB 201 NKPPSSDIIAPPVCGNVEEGERDCGSPEDCOSACCNATCKFKGAGACRAKADDEC 260
QY 39 DLATLATQCSYQCPGTALDDGVTDVFDPSAQCVCCKRNFYNGGSP 86
DB 261 DLPELCTGQSAACP-----TDTFQENKSGCK-KQGYCYNGKCP 298

RESULT 15

P70570 PRELIMINARY; PRT: 1725 AA.

ID P70570;
AC P70570;
DT 01-FEB-1997 (TREMblrel. 02, Created)
DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE LAMININ-5 ALPHA 3 CHAIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Baker S.E., Hopkinson S.B., Fitchmun M., Andreason G.L., Frasier F.,
RA Plopper G., Quaranta V., Jones J.C.R.;
RA J. Cell Sci. 109:0-0(0).
DR EMBL: U61261; AAB17053.1; -;
DR HSSP: P02468; ITLE.
DR INTERPRO: IPR000561; -;

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:53:02 ; Search time 116.78 Seconds
(without alignments)
105.385 Million cell updates/sec

Title: US-09-196-161d-10
Perfect score: 560
Sequence: 1 GSAQGEANGNOFPFAANNAR.....PQGEAFGLQVFAAGAAAGCI 105

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_15:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	544	97.1	394	5 Q27208	Q27208 Ichthyophth
2	544	97.1	442	5 Q9XZ62	Q9XZ62 Ichthyophth
3	84.5	15.1	1737	4 Q75097	Q75097 homo sapien
4	81	14.5	1551	5 Q9NGV4	Q9NGV4 drosophila
5	81	14.5	3396	5 Q9VM55	Q9VM55 drosophila
6	79.5	14.2	874	11 Q9QYPO	Q9QYPO rattus norv
7	79.5	14.0	1019	5 Q9N440	Q9N440 caenorhabdi
8	78.5	14.0	560	5 Q9U013	Q9U013 giardia lam
9	78	13.9	1725	11 P70570	P70570 rattus norv
10	77	13.8	504	5 Q9XZ88	Q9XZ88 leishmania
11	76.5	13.7	400	13 Q9PR47	Q9PR47 atriclastis
12	76	13.6	600	13 Q9PVK7	Q9PVK7 naja naja
13	76	13.6	1372	5 P91526	P91526 caenorhabdi
14	75.5	13.5	1696	5 Q9NJ15	Q9NJ15 branchiosto
15	75	13.4	423	5 Q9U697	Q9U697 tetrahymena
16	75	13.4	439	5 Q94827	Q94827 tetrahymena
17	75	13.4	1086	10 Q9L113	Q9L113 zea mays (m
18	74.5	13.3	2014	5 Q22774	Q22774 caenorhabdi
19	74.5	13.3	2972	5 P90891	P90891 caenorhabdi

20	74	13.2	594	5 Q24970	Q24970 giardia lam
21	74	13.2	1546	4 Q9NS27	Q9NS27 homo sapien
22	74	13.2	1551	4 Q75445	Q75445 homo sapien
23	73.5	13.1	620	13 Q42138	Q42138 agkistrodon
24	73	13.0	187	2 Q07394	Q07394 rhodobacter
25	72.5	12.9	167	5 Q24991	Q24991 giardia lam
26	72.5	12.9	398	1 Q00388	Q00388 methanococ
27	72.5	12.9	401	5 Q9NH87	Q9NH87 giardia lam
28	72	12.9	468	4 Q00220	Q00220 homo sapien
29	72	12.9	598	5 Q9N848	Q9N848 leishmania
30	72	12.9	709	5 Q5XZ99	Q5XZ99 leishmania
31	72	12.9	1792	13 Q57484	Q57484 gallus gall
32	71.5	12.8	498	11 Q60846	Q60846 mus musculu
33	71	12.7	314	4 Q14293	Q14293 homo sapien
34	71	12.7	597	5 Q07317	Q07317 giardia int
35	71	12.7	962	4 Q9Y4H0	Q9Y4H0 homo sapien
36	71	12.7	969	4 Q9UEC7	Q9UEC7 homo sapien
37	71	12.7	975	4 Q9Y4G9	Q9Y4G9 homo sapien
38	71	12.7	975	4 Q9UEJ9	Q9UEJ9 homo sapien
39	71	12.7	1310	2 Q59426	Q59426 escherichia
40	71	12.7	2207	5 Q9U0V2	Q9U0V2 leishmania
41	71	12.7	3110	4 Q93022	Q93022 homo sapien
42	70.5	12.6	531	5 Q9VWZ8	Q9VWZ8 drosophila
43	70	12.5	114	10 Q22484	Q22484 oryza sativ
44	70	12.5	388	10 Q9MA42	Q9MA42 arabidopsis
45	70	12.5	424	5 Q9XZ41	Q9XZ41 ancylostoma

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	394 AA.
Q27208	Q27208	Q27208	Q27208	Q27208
AC	Q27208	Q27208	Q27208	Q27208
DT	01-NOV-1996 (TRENBLER, 01, Created)	01-NOV-1996 (TRENBLER, 01, Created)	01-NOV-1996 (TRENBLER, 01, Created)	01-NOV-1996 (TRENBLER, 01, Created)
DT	01-NOV-1998 (TRENBLER, 08, Last sequence update)	01-NOV-1998 (TRENBLER, 08, Last sequence update)	01-NOV-1998 (TRENBLER, 08, Last sequence update)	01-NOV-1998 (TRENBLER, 08, Last sequence update)
DT	01-JUN-2000 (TRENBLER, 14, Last annotation update)	01-JUN-2000 (TRENBLER, 14, Last annotation update)	01-JUN-2000 (TRENBLER, 14, Last annotation update)	01-JUN-2000 (TRENBLER, 14, Last annotation update)
DE	IMMOBILIZATION ANTIGEN PRECURSOR (FRAGMENT)	IMMOBILIZATION ANTIGEN PRECURSOR (FRAGMENT)	IMMOBILIZATION ANTIGEN PRECURSOR (FRAGMENT)	IMMOBILIZATION ANTIGEN PRECURSOR (FRAGMENT)
OS	Ichthyophthirius multifiliis.	Ichthyophthirius multifiliis.	Ichthyophthirius multifiliis.	Ichthyophthirius multifiliis.
OC	Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;	Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;	Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;	Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC	Opisthokonta; Ichthyophthirius.	Opisthokonta; Ichthyophthirius.	Opisthokonta; Ichthyophthirius.	Opisthokonta; Ichthyophthirius.
OX	NCBI_TaxID=5932;	NCBI_TaxID=5932;	NCBI_TaxID=5932;	NCBI_TaxID=5932;
RN	[1]	[1]	[1]	[1]
RP	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.
RC	STRAIN=GEORGIA;	STRAIN=GEORGIA;	STRAIN=GEORGIA;	STRAIN=GEORGIA;
RC	MEDLINE=92335298; PubMed=1631132;	MEDLINE=92335298; PubMed=1631132;	MEDLINE=92335298; PubMed=1631132;	MEDLINE=92335298; PubMed=1631132;
RX	Clark T.G., McGraw R.A., Dickerson H.W.;	Clark T.G., McGraw R.A., Dickerson H.W.;	Clark T.G., McGraw R.A., Dickerson H.W.;	Clark T.G., McGraw R.A., Dickerson H.W.;
RA	"Developmental expression of surface antigen genes in the parasitic	"Developmental expression of surface antigen genes in the parasitic	"Developmental expression of surface antigen genes in the parasitic	"Developmental expression of surface antigen genes in the parasitic
RT	ciliate Ichthyophthirius multifiliis.";	ciliate Ichthyophthirius multifiliis.";	ciliate Ichthyophthirius multifiliis.";	ciliate Ichthyophthirius multifiliis.";
RL	Proc. Natl. Acad. Sci. U.S.A. 89:6363-6367(1992).	Proc. Natl. Acad. Sci. U.S.A. 89:6363-6367(1992).	Proc. Natl. Acad. Sci. U.S.A. 89:6363-6367(1992).	Proc. Natl. Acad. Sci. U.S.A. 89:6363-6367(1992).
RN	[2]	[2]	[2]	[2]
RP	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.
RC	STRAIN=GEORGIA;	STRAIN=GEORGIA;	STRAIN=GEORGIA;	STRAIN=GEORGIA;
RC	MEDLINE=93020590; PubMed=1383510;	MEDLINE=93020590; PubMed=1383510;	MEDLINE=93020590; PubMed=1383510;	MEDLINE=93020590; PubMed=1383510;
RA	Liu T.L., Dickerson H.W.;	Liu T.L., Dickerson H.W.;	Liu T.L., Dickerson H.W.;	Liu T.L., Dickerson H.W.;
RT	"Purification and partial characterization of immobilization antigens	"Purification and partial characterization of immobilization antigens	"Purification and partial characterization of immobilization antigens	"Purification and partial characterization of immobilization antigens
RL	from Ichthyophthirius multifiliis.";	from Ichthyophthirius multifiliis.";	from Ichthyophthirius multifiliis.";	from Ichthyophthirius multifiliis.";
RN	J. Protozool. 39:457-463(1992).	J. Protozool. 39:457-463(1992).	J. Protozool. 39:457-463(1992).	J. Protozool. 39:457-463(1992).
RN	[3]	[3]	[3]	[3]
RP	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.
RC	STRAIN=GEORGIA;	STRAIN=GEORGIA;	STRAIN=GEORGIA;	STRAIN=GEORGIA;
RC	Clark T.;	Clark T.;	Clark T.;	Clark T.;
RA	Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.	Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.	Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.	Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
RN	[4]	[4]	[4]	[4]
RP	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.
RC	STRAIN=GEORGIA;	STRAIN=GEORGIA;	STRAIN=GEORGIA;	STRAIN=GEORGIA;
RA	Clark T.;	Clark T.;	Clark T.;	Clark T.;
RL	Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.	Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.	Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.	Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR	EMBL: M92907; AAC36158.1; -.	EMBL: M92907; AAC36158.1; -.	EMBL: M92907; AAC36158.1; -.	EMBL: M92907; AAC36158.1; -.
KW	Signal.	Signal.	Signal.	Signal.
FT	NON_TER	NON_TER	NON_TER	NON_TER
FT	SIGNAL	SIGNAL	SIGNAL	SIGNAL

```

SQ SEQUENCE 394 AA; 39495 MW; 3013C2B2BEFDB62 CRC64;

Query Match
Best Local Similarity 97.1%; Score 544; DB 5; Length 394;
Matches 101; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 GSAOGEANGNOPFAANNAAARGICVPCQINRVSGTNGADLTATLTCSTGCTGTALDDG 60
   :|||||
Db 36 GAAOGEANGNOPFAANNAAARGICVPCQINRVSGTNGADLTATLTCSTGCTGTALDDG 95
   :|||||

OY 61 VTDFEDSSAACVCKRPFYNYNGSSPGGEARGLOVFAAGAAACI 105
   :|||||
Db 96 VTDFEDSSAACVCKRPFYNYNGSSPGGEARGVGFVFAAGAAAGV 140
   :|||||

RESULT 2
O9XZG2 PRELIMINARY; PRT; 442 AA.
AC O9XZG2;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE IMMOBILIZATION ANTIGEN PRECURSOR.
GN TAG48.
OS Ichthyophthirius multifiliis.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Ophryoglenina; Ichthyophthirius.
OX NCBI_TaxID=5932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GI;
RX MEDLINE=99196987; PubMed=10095108;
RA Clark T.G., Lin T.L., Jackwood D.A., Sherrill J., Lin Y.,
  Dickerson H.W.;
RT "The gene for an abundant parasite coat protein predicts tandemly
  repetitive metal binding domains.";
RL Gene 229:91-100(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GI;
RA Gaertig J., Gao Y., Tishgarten T., Clark T.G., Dickerson H.W.;
RT "Surface display of a parasite antigen in the ciliate Tetrahymena
  thermophila.";
RL Nat. Biotechnol. 0:0-0(1999).
DR EMBL; AF140273; AAD31283.1; -.
KM Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 442 IMMOBILIZATION ANTIGEN.
FT SEQUENCE 442 AA; 45025 MW; 52658F3F65D27AFA CRC64;

Query Match
Best Local Similarity 97.1%; Score 544; DB 5; Length 442;
Matches 101; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 GSAOGEANGNOPFAANNAAARGICVPCQINRVSGTNGADLTATLTCSTGCTGTALDDG 60
   :|||||
Db 56 GAAOGEANGNOPFAANNAAARGICVPCQINRVSGTNGADLTATLTCSTGCTGTALDDG 115
   :|||||

OY 61 VTDFEDSSAACVCKRPFYNYNGSSPGGEARGLOVFAAGAAACI 105
   :|||||
Db 116 VTDFEDSSAACVCKRPFYNYNGSSPGGEARGVGFVFAAGAAAGV 160
   :|||||

RESULT 3
O75097 PRELIMINARY; PRT; 1737 AA.
AC O75097;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE MEGF8 (FRAGMENT).

```

```

GN MEGF8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=98360089; PubMed=9693030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple EGF-
  like motifs by motif-trap screening.";
RL Genomics 51:27-34(1998).
DR EMBL; AB011541; BAA32469.1; -.
DR HSSP; P02468; TITLE.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000859; -.
DR INTERPRO; IPR001881; -.
DR PFAM; PF00053; laminin_EGF_2.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_4.
DR PROSITE; PS01180; CUB_1.
DR PROSITE; PS01186; EGF_2; 4.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; UNKNOWN_4.
FT NON_TER 1 1
FT SEQUENCE 1737 AA; 184680 MW; FD3152E415193159 CRC64;

Query Match
Best Local Similarity 15.1%; Score 84.5; DB 4; Length 1737;
Matches 25; Conservative 4; Mismatches 28; Indels 17; Gaps 3;

OY 21 GICVPCQINRVSGTNGADLTATLTCSTGCTGTALDDGVDFEDSSAACVCK 74
   :|||||
Db 1267 GKCTKQCQN-----GHADTCNEQDGTGCPQNNMTETGCGSSPSDRDRCKYQCA 1318
   :|||

OY 75 CKPFYNYNGSSPG 88
   :|||
Db 1319 CRESFH---GSPLG 1329
   :|||

RESULT 4
O9NGV4 PRELIMINARY; PRT; 1551 AA.
AC O9NGV4;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE SP1070.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ENDOPLASMIC RETICULUM;
RA Serano T.L., Pendleton J.D., Rubin G.M.;
RT "A reverse genetic screen for genes involved in Drosophila
  development.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF239608; AAF63500.1; -.
FT SEQUENCE 1551 AA; 167816 MW; A97EA229E9384F31 CRC64;

Query Match
Best Local Similarity 14.5%; Score 81; DB 5; Length 1551;
Matches 28; Conservative 3; Mismatches 43; Indels 20; Gaps 3;

OY 1 GSAOGEANGNOPFAANNAAARGICVPCQINRVSGTNGADLTATLTCSTGCTGTALDDG 60
   :|||||

```

Db 1127 GTYGSEAOLO-----CSKCPYIAGRPVGTATPGARSADCKERCAG----- 1169

QY 61 VTDFDRSAOCVKCKPNFY-NGSGPQGEAPGL 93

Db 1170 --KYFDAQETGLCRSGHGFGYQPNESGFCSELCGL 1201

RESULT 5

Q9VMS5 PRELIMINARY; PRT; 3396 AA.

AC 09VMS5: 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, last sequence update)

DT 01-OCT-2000 (Tremblrel. 15, last annotation update)

DE CG9138 PROTEIN.

OS CG9138.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RA MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abdl J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasly E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jaitani B., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Klumel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacלב J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spraker R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weissstock G.M., Weissenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,

RT "The genome sequence of Drosophila melanogaster."

RL Science 287:2185-2195(2000).

DR EMBL: AEO03615; AAF52472.1; -

DR HSSP: P00740; IEDM

DR FLXBASE: FBgn001879; CG9138.

DR INTERPRO: IPR000152; -

DR INTERPRO: IPR000421; -

DR INTERPRO: IPR000436; -

DR INTERPRO: IPR000561; -

DR INTERPRO: IPR000742; -

DR INTERPRO: IPR000859; -

DR INTERPRO: IPR001092; -

DR INTERPRO: IPR001438; -

DR INTERPRO: IPR001791; -

DR INTERPRO: IPR001881; -

DR INTERPRO: IPR002172; -

DR PFAM: PF000068; EGF; 16.

DR PFAM: PF00057; IGL_recept_a; 1.

DR PFAM: PF00084; sushi; 7.

DR PFAM: PF00431; CUB; 3.

DR PFAM: PF00754; F5_F8_type_C; 1.

DR PRINTS: PR00010; EGFBLD.

DR PROSITE: PS00010; ASX_HYDROXYL; 11.

DR PROSITE: PS00022; EGF_1; 15.

DR PROSITE: PS01180; CUB; 3.

DR PROSITE: PS01186; EGF_2; 13.

DR PROSITE: PS01187; EGF_CA; 7.

DR PROSITE: PS01209; IDLRA_1; 1.

DR PROSITE: PS01285; FA58C_1; 1.

DR PROSITE: PS01285; FA58C_2; 2.

SO SEQUENCE 3396 AA; 369389 MW; E618E9ACEA13E0E5 CRC64;

Query Match 14.5%; Score 81; DB 5; Length 3396;

Best Local Similarity 29.8%; Pred. No. 4.3;

Matches 28; Conservative 3; Mismatches 43; Indels 20; Gaps 3;

QY 1 GSAOGEANGNPFANNARAGICVPCQINRVSGTNAGDLATLATQGSTQCPGTALDDG 60

Db 2968 GTYGSEAOLO-----CSKCPYIAGRPVGTATPGARSADCKERCAG----- 3010

QY 61 VTDFDRSAOCVKCKPNFY-NGSGPQGEAPGL 93

Db 3011 --KYFDAQETGLCRSGHGFGYQPNESGFCSELCGL 3042

RESULT 6

Q9QYPO PRELIMINARY; PRT; 874 AA.

AC 09QYPO: 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, last sequence update)

DT 01-OCT-2000 (Tremblrel. 15, last annotation update)

DE MEGF8 (FRAGMENT).

GN MEGF8.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;

RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.,

RT "Identification of high-molecular-weight proteins with multiple EGF-

RT like motifs by motif-trap screening."

RL Submitted (Feb-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AB011534; BAA86869.1; -

DR HSSP: P35555; IEMN.

DR INTERPRO: IPR000152; -

DR INTERPRO: IPR000561; -

DR INTERPRO: IPR001881; -

DR INTERPRO: IPR002049; -

DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.

DR PROSITE: PS00022; EGF_1; UNKNOWN_1.

DR PROSITE: PS01186; EGF_2; 2.

DR PROSITE: PS01248; LAMININ_Type_EGF; UNKNOWN_2.

FT NON_TER 1

SO SEQUENCE 874 AA; 93764 MW; CB63BDF183B870AB CRC64;

Query Match 14.2%; Score 79.5; DB 11; Length 874;

Best Local Similarity 32.4%; Pred. No. 1.6;

Matches 24; Conservative 4; Mismatches 29; Indels 17; Gaps 3;

QY 21 GICVPCQINRVSGTNAGDLATLATQGSTQCP-----GTALDDGVDFDRSAOCVK 74

Db 404 GKCTKCCOCCN-----GHADTCNEODTGCPCONNTETGVCOCSSPSDRDRCKYKCAK 455
OY 75 CKPNEYNGSGSPG 88
Db 456 CRESFH--GSPFLG 466

RESULT 7
O9NMA40
ID O9NMA40 PRELIMINARY; PRT; 1019 AA.
AC O9NMA40;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE Y73F8A.5 PROTEIN.
GN Y73F8A.5
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peioderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews L.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA none;
RT "genome sequence of the nematode C. elegans: A platform for
investigating biology."
RT Science 282:2012-2018(1998).
RL EMBL; AL132862; CAB70224.1;
SQ SEQUENCE 1019 AA; 112378 MW; 6EB4094722B707E7 CRC64;

Query Match 14.28; Score 79.5; DB 5; Length 1019;
Best Local Similarity 29.98; Pred. No. 1.8;
Matches 26; Conservative 12; Mismatches 26; Indels 23; Gaps 5;

OY 10 NOPFAANNAARGICVPCOINVGSGTNAAGLATLA-TCCSTCOCPTGTALDGVTFEDRS 68
Db 242 NMGOAANO-----PIHIDRI-----DVLGIATFROCTACPGTSSPGG----- 280

OY 69 AAOVCVKCPNEYNGSGPO-GEAPGLQ 94
Db 281 SAECIPCSGFSKSGGCGCRCPESQ 307

RESULT 8
O9U013
ID O9U013 PRELIMINARY; PRT; 560 AA.
AC O9U013;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE VARIANT-SPECIFIC SURFACE PROTEIN H7-1.
GN H7-1.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
OX NCBI_Taxid=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS;
RX MEDLINE=96010591; Pubmed=7581335.
RA Nash T.E., Conrad J.T., Mowatt M.R.;
RT "Giardia lamblia: Identification and characterization of a variant-
specific surface protein gene family."
RL J. Eukaryot. Microbiol. 42:604-609(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GS;
RA Nash T.E., Mowatt M.R., Conrad J.T.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF202776; AAF17594.1;
SQ SEQUENCE 560 AA; 57022 MW; 6816D5713249BA8 CRC64;

Query Match 14.08; Score 78.5; DB 5; Length 560;
Best Local Similarity 27.68; Pred. No. 1.3;
Matches 35; Conservative 8; Mismatches 47; Indels 37; Gaps 7;

OY 14 AANNAARG---ICVPCO-----NRVGS-----GTNAGDLATLATGCGTCTGALD 58
Db 224 ACNTRANGIDKCAECSLITPASRAGALLITCTKCSINS--LSPKDACLTSCFAGTYET 281

OY 59 DGVTDFV-----DRSAQCVKCPNEYNGSGPO-----GEAPG--LOYFAA 98
Db 282 GSPKVCPCPHTSCAGCKDDMTAASCTACYPGSVLSYSGSDTKGTCTIAECTGKYLENAD 341

OY 99 GAANAAGI 105
Db 342 GOCCTASI 348

RESULT 9
P70570
ID P70570 PRELIMINARY; PRT; 1725 AA.
AC P70570;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE LAMININ-5 ALPHA 3 CHAIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Baker S.E., Hopkinson S.B., Fitchmun M., Andreason G.L., Frasier F.,
RA Plopper G., Quaranta V., Jones J.C.R.;
RL J. Cell Sci. 109:0-0(0).
DR EMBL; U61261; AAB17053.1;
DR HSSP; P02468; IITLE
DR INTERPRO; IPR000561;
DR INTERPRO; IPR001791;
DR INTERPRO; IPR002049;
DR PFAM; PF00053; laminin_EGF_2.
DR PFAM; PF00054; laminin_G_3.
DR PROSITE; PS00023; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01248; LAMININ_TYPE_EGF_2.
KW Glycoprotein.
SQ SEQUENCE 1725 AA; 190392 MW; 02EE43B33E72B0FB CRC64;

Query Match 13.98; Score 78; DB 11; Length 1725;
Best Local Similarity 28.78; Pred. No. 4.6;
Matches 25; Conservative 9; Mismatches 31; Indels 22; Gaps 5;

OY 21 GICVPCOINRVGS-----GTNAGDLATLATGCGTCTGALDGVTFEDR 67
Db 92 GVCINCOHNTAGEHCERCKKRYGSAIHGSCRVC--PCPHNTPATGAVDGAVRACK 149

OY 68 ---SAOVCVKCPNEYNYN---GGSPQ 87
Db 150 PGYTGACERCACPGYFGNPKFGGSCQ 176

RESULT 10
O9XZ8
ID O9XZ8 PRELIMINARY; PRT; 504 AA.
AC O9XZ8;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE L302.1 PROTEIN.

GN L302.1
OS Leishmania major.
OC Eukaryota; Euzoenzoa; Kinetoplastida; Trypanosomatidae; Leishmania
OX NCBI_TaxId=9664;
RN [1]
RP SEQUENCE FROM N.A.
RA Murphy L., Quall M., Lawson D., Harris D., Rajadream M., Ivens A.,
RA Barrell B.;
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AL031909; CAB42343.1; -
DR INTERPRO: IPR000564; -
DR INTERPRO: IPR002049; -
DR PROSITE: PS00197; 2FE2S_FERRDOXIN; 3.
DR PROSITE: PS01248; LAMININ_TYPE_ECG; UNKNOWN.1
SO SEQUENCE 504 AA; 52833 MW; B75450E5FE59263 CRC64;

	Query Match	13.8%	Score 77;	DB 5;	length 504;
	Best Local Similarity	32.4%;	Pred. No. 1.6;		
	Matches 22;	Conservative	8;	Mismatches	26; Indels 12; Gaps
QY	21	GICVPCOINRGGSTNAGDLATLATCOSTCPPTALDDGV-----TDVF--DRSAACG	72		
		: : :			
Dd	253	GVCYLPOQL---CTIDANCASCPSDAGTC-TCCANGYGVLADCAVRCQEPNCFCSDSDANKC	308		
		: : :			
QY	73	VKCRPNFY	80		
		: :			
Dd	309	TCCAPNTY	316		

```

RESULT 11
Q9PT47
ID Q9PT47 PRELIMINARY; PRT; 400 AA

```

DT	01-MAY-2000	(TREMBLER, 13, Created)
DT	01-MAY-2000	(TREMBLER, 13, last sequence update)
DT	01-OCT-2000	(TREMBLER, 15, last annotation update)
DE	METALLOPROTEINASE PRECURSOR (FRAGMENT).	
OS	Atractaspis engadadensis (Israeli burrowing asp).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Lepidosaurs; Squamata; Scleroglossa; Serpentes; Colubroidea;	
OC	Atractaspididae; Atractaspis.	
OX	NCBI_TaxId=8600;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=VENOM GLAND;	
RA	Duncan F.;	
RL	Submitted (SEP-1989) to the EMBL/Genbank/DBJ databases.	
DR	EMBL; AF186370; AAFO1042.1; -;	
DR	HSSP; P17494; 1KST;	
DR	INTERPRO; IPR000130; -;	
DR	INTERPRO; IPR001590; -;	
DR	INTERPRO; IPR001762; -;	
DR	PFAM; PF00200; disintegrin; 2.	
DR	PF01421; Reptolysin; 1.	
DR	PRINTS; PRO0289; DISINTEGRIN.	
DR	PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.	
FT	NON_TER	1
FT	CHAIN	1 203
FT	CHAIN	204
FT	CHAIN	280
SQ	SEQUENCE	400 AA; 44790 MW; B562FAFD2EB79468 CRC64;

Query Match	13.7%	Score 76.5;	DB 13;	Length 400;
Best Local Similarity	23.18%	Pred. No. 1.5;		
Matches 25; Conservative	10;	Mismatches 32;	Indels 41;	Gaps 4;

[illegible]

Db 261 DLPELCTGQSAECP-----TPTFQENGKSCQK-KQGYCYNGKCP 298

RESULT 1

ID	Q9PVK7	PRELIMINARY;	PRT;	600 AA
----	--------	--------------	------	--------

DT 01-MAY-2000 (TEMBLrel, 13, created)
DT 01-MAY-2000 (TEMBLrel, 13, last sequence update)
DT 01-OCT-2000 (TEMBLrel, 15, last annotation update)
DE COBRIN PRECURSOR.
OS Najia najia (Indian cobra)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Elapinae; Najia
NCBI_TaxID=35670;

RP SEQUENCE FROM N.A.
RC STRAIN=KAOUTHIA; TISSUE=VENOM GLAND;
RA Bambaï B., Bredehorst R., Vogel C.-W.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases
EMBL: AB063190.1; GenBank: AF070693.1.

DR HSBV; F1/494; INSI; -
DR INTERPRO: IPR000130; -
DR INTERPRO: IPR001590; -
DR INTERPRO: IPR001762; -
DR INTERPRO: IPR002870; -
DR PFM; PF00200; distalysin, 1.
DR PFM; PF01421; distalysin, 1.
DR PFM; PF01562; pep_M12b_propep, 1.
DR PRINTS: PFO0289; DISTINTEGRIN.
DR POSITIVE: PS00142; ZINC_PROTEASE; UNKNOWN_1
DR Pfam01421; distalysin, 1

FT	SIGNAL	1	6	POTENTIAL.
FT	CHAIN	180	600	COBRIN.
SQ	SEQUENCE	600 AA;	67661 MW;	C7C0D45EBC694290 CRC64;

Query Match	13.6%	Score	76	DB	13	Length	600
Best Local Similarity	30.6%	Pred. No.	2.5				
Matches	26	Conservative	8	Mismatches	27	Indels	24
				Gaps			5

DQ 23 CVCPCQINRGSGTNG---DLATLAIQCSTGCPPTGTALDDGVTVYEDRSAAQCCKKPN 78
| | : | | | : ||
Db 442 CECKCFKGAGAECCRAAKDCDLPCLCTGQSACGP-----TDVFQRNG---LPCQNN 489

```
QY 79 FY-YNGSP-----QGEAPGLQV 95
      | ||| |
      ||::|
Db 490 GCYNGKCPIMTNCIALRGPVKV 514
```

RESULT 13

DP 01-MAY-1997 (TREMBLER. 03, Created)
 DT 01-MAY-1997 (TREMBLER. 03, Last sequence update)
 DT 01-OCT-2000 (TREMBLER. 15, Last annotation update)
 DE SMILITARITY TO MULTIPLE EGF-LIKE DOMAINS.
 GN W02C12.1..
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 NCBI_TaxID=6239;

RC STRAIN-BRISTOL, N2: 1
EX MEDLINE-94150718; PubMed-7906399;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M., Coulson A.
RA Bonfield J., Burton J., Connell M., Copsey J., Cooper J.,
RA Craxton M., Dear S., Du Z., Duplin R., Favell A., Fulton L.,
RA Gardner A., Green P., Hawkins J., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,

RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Riffen L., Roopra A., Saunders D., Showkeen R.,
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierly-Wieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sprat J., Woldman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Murray J., Woldmann P.,
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U80815; AAB3795.1; -;
 DR HSP: Q12784; IIRE.
 DR INTERPRO: IPR000152; -;
 DR INTERPRO: IPR000561; -;
 DR INTERPRO: IPR000742; -;
 DR INTERPRO: IPR001438; -;
 DR INTERPRO: IPR001881; -;
 DR PFAM: PF00008; EGF_12.
 DR PRINTS: PR00010; EGFBL00D.
 DR PROSITE: PS00010; ASX_HYDROXYL; 5.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_11.
 DR PROSITE: PS01186; EGF_2; 8.
 DR PROSITE: PS01187; EGF_CA; 2.
 KW Glycoprotein; EGF-like domain.
 SQ SEQUENCE 1372 AA; 150561 MW; 0537FA79BA0576C5 CRC64;

Query Match 13.6%; Score 76; DB 5; Length 1372;
 Best Local Similarity 21.1%; Pred. No. 5.9;
 Matches 28; Conservative 15; Mismatches 50; Indels 40; Gaps 5;

QY 2 SAGEANGNOPFAANNAARGICVPCQINRVGS-----GTNAGDLATLATQCS 48
 Db 1035 TSEBCEDECDPGBOLASVSGCPCQIGTYSRGENKCVACPGTTEATMSRREDCN 1094
 QY 49 TO-----CPTGALDGVTVDFDRSAACVCKKPNFYNGSGPGEAP 91
 Db 1095 TPCKRPGQFLVKETKNCQFCRGTFFON-----EEQESTCKLCAPD--HTTAAP-GATA 1144
 QY 92 GLQVFAGAAAAG 104
 Db 1145 ESQCFSTNOCATG 1157

RESULT 14
 O9N15
 ID O9N15 PRELIMINARY; PRT; 1696 AA.
 AC O9N15;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE PROTEIN CONVERTASE APC6B ISOFORM.
 GN PC6.
 OS Branchiostoma californiensis (California lancelet) (Amphioxus).
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
 OC Branchiostoma.
 OX NCBI_TaxID=7738;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Olive A.A. Jr., Chan S.J., Steiner D.F.;
 RT "Evolution of the prohormone convertases: identification of a
 RT homologue of pc6 in the protochordate amphioxus.";
 RL Biochim. Biophys. Acta 1477:338-348(2000).
 DR EMBL: AF184616; AAF26301.1; -;
 SQ SEQUENCE 1696 AA; 188409 MW; 281CB1784257CDB CRC64;

Query Match 13.5%; Score 75.5; DB 5; Length 1696;
 Best Local Similarity 29.4%; Pred. No. 8.4;
 Matches 20; Conservative 3; Mismatches 28; Indels 17; Gaps 2;
 QY 23 CVPCQINRVSGTNAGDLATLATQCSPTGTALDDGVTVDFRSA-----QCVK 74
 Db 1303 CTAC-----NDGFILTDASSCEACCPGQFLHHDQDCSCHRECKTCDDGPHNDCLS 1353
 QY 75 CKPNFYNN 82
 Db 1354 CQPGSYLN 1361

RESULT 15
 O9U697
 ID O9U697 PRELIMINARY; PRT; 423 AA.
 AC O9U697;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE SERH3 IMMOBILIZATION ANTIGEN.
 GN SERH3.
 OS Tetrahymena thermophila.
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
 OC Tetrahymena; Tetrahymena.
 OX NCBI_TaxID=5911;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Clark T.G., Cheng G.;
 RT "Serh3 Immobilization Antigen from Tetrahymena thermophila.";
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF190647; AAF06326.1; -;
 SQ SEQUENCE 423 AA; 41026 MW; B945BCB2E491E6B6 CRC64;

Query Match 13.4%; Score 75; DB 5; Length 423;
 Best Local Similarity 31.1%; Pred. No. 2.3;
 Matches 33; Conservative 9; Mismatches 46; Indels 18; Gaps 5;

QY 2 SAGEANGNOPFAANNAARGICVPCQINRVSGTNA---GDLATLATQCSPTGT 55
 Db 22 SAGQANGTGVAACTDCASVCGVP---TVAGTGTACSWSSSTLTCTVDTCTLTGT 78
 QY 56 ALDDGVTVDFDRSAACVCKKPNFYNGSGPGEAPGLQVFAAGAA 101
 Db 79 V--TGITNLNDQFCTSCGSTNTYANGA-----GTACVAAASAS 115

Search completed: March 6, 2001, 12:53:04
 Job time: 304 sec

Tri Mar 9 15:47:44 2001

us-09-196-161d-10.rspt

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:53:53 ; Search time 57.76 Seconds
(without alignments)
32.644 Million cell updates/sec

Title: US-09-196-161d-10

Perfect score: 560

Sequence: 1. GSAOGFANGNOPFANNNAAR.....POGEAPGLQVFAAGAAAGI 105

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	72	12.9	335	3	US-08-815-469-6
2	71	12.7	119	2	US-08-219-2378-3
3	71	12.7	219	3	US-08-974-022-45
4	71	12.7	314	1	US-08-444-231-19
5	71	12.7	314	1	US-08-152-443A-19
6	71	12.7	314	4	PCT-US95-17083-4
7	71	12.7	335	2	US-08-219-2378-2
8	71	12.7	335	2	US-08-409-338-1
9	71	12.7	335	4	PCT-US95-17083-2
10	71	12.7	969	2	US-08-284-941-2
11	71	12.7	969	2	US-08-447-642-2
12	71	12.7	969	4	PCT-US93-02147A-2
13	71	12.7	3111	2	US-08-460-309-4
14	71	12.7	3111	2	US-08-125-077-4
15	70.5	12.6	219	1	US-08-152-019A-32
16	70	12.5	314	2	US-08-460-309-19
17	70	12.5	314	2	US-08-125-077-19
18	70	12.5	424	1	US-08-419-414-2
19	69	12.3	169	2	US-08-460-309-20
20	69	12.3	169	2	US-08-125-077-20
21	69	12.3	1111	1	US-08-317-5508-15
22	69	12.3	1111	3	US-08-800-593-15
23	69	12.3	1111	1	US-08-317-450B-13
24	69	12.3	1193	3	US-08-800-593-13
25	68.5	12.2	341	2	US-08-209-521-11
26	68.5	12.2	610	3	US-08-365-470-3
27	68.5	12.2	610	3	US-09-209-668-19
28	68.5	12.2	610	5	5217870-2

29	67	12.0	156	3	US-08-600-982-30	Sequence 30, Appl
30	67	12.0	156	4	PCT-US94-10261A-30	Sequence 30, Appl
31	67	12.0	1713	3	US-08-600-982-24	Sequence 24, Appl
32	67	12.0	1713	4	PCT-US94-10261A-24	Sequence 24, Appl
33	66.5	11.9	197	2	US-08-505-606-1	Sequence 1, Appl
34	66	11.8	111	1	US-08-288-728-4	Sequence 4, Appl
35	65	11.6	3075	2	US-08-460-309-5	Sequence 5, Appl
36	65	11.6	3075	2	US-08-125-077-5	Sequence 5, Appl
37	64.5	11.5	484	2	US-08-252-493C-9	Sequence 9, Appl
38	64.5	11.5	484	3	US-09-276-197-9	Sequence 9, Appl
39	64.5	11.5	846	2	US-07-728-215-33	Sequence 3, Appl
40	62.5	11.2	225	1	US-08-152-019A-33	Sequence 3, Appl
41	62	11.1	870	2	US-09-010-928B-2	Sequence 2, Appl
42	61.5	11.0	211	1	US-08-276-852-34	Sequence 34, Appl
43	61.5	11.0	211	1	US-08-133-011-16	Sequence 16, Appl
44	61.5	11.0	211	1	US-08-322-730A-16	Sequence 16, Appl
45	61.5	11.0	211	1	US-08-387-874-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-08-815-469-6
Sequence 6, Application US/08815469
Patent No. 6153402
GENERAL INFORMATION:
APPLICANT: Yu, Guo-Liang
APPLICANT: Ni, Jian
APPLICANT: Dixit, Vishva
APPLICANT: Gentz, Reiner L.
APPLICANT: Dillon, Patrick J.
TITLE OF INVENTION: Death Domain Containing Receptors
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox, P.L.L.C.
STREET: 1100 New York Ave., NW, Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/815,469
FILING DATE: HERewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: No. 6153402 Yet Assigned
FILING DATE: 06-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,711
FILING DATE: 17-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/013,285
FILING DATE: 12-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0310003/EKS/KRM
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 335 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein

US-08-815-469-6

Query Match	12.9%;	Score 72;	DB 3;	Length 335;
Best Local Similarity	28.6%;	Pred. NO. 4.4;		
Matches 22;	Conservative 10;	Mismatches 27;	Indels 18;	Gaps 4;

```

OY      8 NGNPPFAANNARGICVPCQINRVSGSINAGDLATLATQC--STQCPGTALDDGVIDVF 65
      ||::| |||| | |::| | |::|
Db     76 NGDEP-----DCVPCQ----EKKEYITDKAHFSSKRCRCRLCDGSHGLEVEIN--C 119

```

```
Qy      66  DRSAAQCVKCKPNFFYYN  82
          | : ||||| : |
Db      120  TRTQNTKCRCKPNFFQON  136
```

RESULT 2

; Sequence 3, Application US/08219237B
; Patent No. 5874546
GENERAL INFORMATION:

```

; MOLECULE TYPE: protein
US-08-219-237B-3

```

Query Match	12.7%	Score 71:	DB 2:	Length 119:
Best Local Similarity	28.6%	Pred. NO	1.7:	
Matches 22:	Conservative 10:	Mismatches 27:	Indels 18:	Gaps 4:

```
QY      66  DRSAAQCCKCPNFYYN  82
      .  | : |||| : |
Db      74  TRTQNTKCRCKPNFFCN  90
```

RESULT 3

US-08-974-022-45
; Sequence 45, Application US/08974022

Query Match Similarity	12.7%	Score	71	DB	3	Length	219
Best Local Similarity	28.6%	Pred. No.	3.4				
Matches	22	Conservative	10	Mismatches	27	Indels	18
						Gaps	4

```

QY      8 NGNQPFAANNARGLCVPCQINRVSGSTNAGDIATLATOC--STOCPGTGALDDGVTVDF 65
      .  ||::|      ||||      |      |      |      |      |      |
Db      76 NGDEP-----DCVPCQ-----ECKEYITDAHFSSKCRRCRLDEGHGLEVEIN--C 119

```

```
QY      66  DRSAAQCVKCKPNFYNN  82
          | : : ||||| : |
Db      120 TRTQNTKCRCKPNFEFCN  136
```

RESULT 4

US-08-444-231-19
; Sequence 19, Application US/08444231
; Patent No. 5652210

STATE: California
COUNTRY: USA
ZIP: 94304-1018

RESULT 3

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5500
; TELEFAX: (415) 494-0792
; TELEEX: 706141
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 314 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-152-443A-19

Query Match      12.7%; Score 71; DB 1; Length 314;
Best Local Similarity 28.6%; Pred. No. 5.2;
Matches 22; Conservative 10; Mismatches 18; Gaps 4.

OY      8 NGNOFANMNAARGICVCQOINRVSGSNAGGLATLATQC--STQCPSGTALDDGCTDYF 65
          |||||          | | :||| | | | | | | | | | | | | | | | | |
Db      76 NGDEP-----DVCVPCQ---EGKEYTIDKAHFSSKCRRCRLCDESHGLEVEIN--C 119
          |||||          | | :||| | | | | | | | | | | | | | | | | |

OY      66 DRSAAQCVCCKPNFYNN 82
          | : ||||| : |
Db     120 TRTQNTKCRCKPNFRCN 136

```

```

RESULT      6
PCT-US95-17083-4
: Sequence 4, Application PC/TUS9517083
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: SECRETED HUMAN FAS ANTIGEN
: NUMBER OF SEQUENCES: 16
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/17083
: FILING DATE: CONCURRENTLY HEREMITH
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: USSN 08/371,263
: FILING DATE: 23-DEC-1994
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 314 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
PCT-US95-17083-4

Query Match      12.7%; Score 71; DB 4; Length 314;
Best Local Similarity 28.6%; Pred. No. 5.2;
Matches 22; Conservative 10; Mismatches 27; Indels 18; Gaps 4;

QY      8 NGNDFANNANMARGICVPCQINRWGSGTNAGDLATATQC--STQCPGTALDDGVTDF 65
      |||:|      |||||      |      |      |      |      |
Db      76 NGDEP-----DCVFCQ-----EGKEVYTDKAHFSSKRCRCRLDGGHGLEVEIN--C 119
      |      |      |      |      |      |      |      |
QY      66 DRSAAQCVKCKRPNTYIN 82
      |      |      |      |      |      |      |      |
Db      120 TRQNTKCRCKRPNEFCN 136

RESULT      7
US-08-219-237B-2
: Sequence 2, Application US/08219237B
: Patent No. 5874546
: GENERAL INFORMATION:
: APPLICANT: NAGATA, Shigekazu
: APPLICANT: ITOH, Naoto
: APPLICANT: YONEHARA, Shin
: TITLE OF INVENTION: DNA Coding for Human Cell Surface Antigen
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: James W. Hellwege

```

FOR I WANT: FALCENCLIN RELEAS

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:50:59 ; Search time 70.34 Seconds
(without alignments)
101.359 Million cell updates/sec

Title: US-09-196-161d-10
Perfect score: 560
Sequence: 1 GSAOGEANGNOPFAANNAAR.....PGEAPGLQVFAAGAAAGI 105

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 segs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891 1

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	544	97.1	395	2 A46031	Immobilization sur
2	94.5	16.9	677	2 C42125	trophozoite cystel
3	84.5	15.1	713	2 A35502	major surface-labe
4	84.5	15.1	1737	2 T00209	MEGF8 protein - hu
5	82	14.6	3635	2 T10053	laminin alpha 5 ch
6	76	13.6	1372	2 T25933	hypothetical prote
7	75	13.4	439	2 A36385	surface antigen se
8	74.5	13.3	573	2 D83130	probable iron-sulf
9	74.5	13.3	2014	2 T21560	hypothetical prote
10	74.5	13.3	2824	2 T22759	hypothetical prote
11	73	13.0	187	2 S39694	probable iron-sulf
12	72.5	12.9	398	1 S24802	polyferredoxin 6x2
13	71	12.7	314	2 I37383	FAS soluble protei
14	71	12.7	335	2 A40036	apoptosis-mediatio
15	71	12.7	596	2 A45664	variant-specific s
16	71	12.7	969	1 A39490	subtilisin-like pr
17	71	12.7	975	2 JC5570	subtilisin-like pr
18	71	12.6	1310	1 T53597	proline dehydrogen
19	70.5	12.6	1607	1 MMSB2	laminin gamma-1 ch
20	70	12.5	114	2 T02043	lipid transfer pro
21	70	12.5	1751	1 MMHUMH	laminin alpha-2 ch
22	70	12.5	3084	1 MMSA	laminin alpha-1 ch
23	69.5	12.4	3712	2 S18253	laminin alpha-1 ch
24	69.5	12.4	289	2 T25682	hypothetical prote
25	69.5	12.4	474	2 T27297	hypothetical prote
26	69.5	12.4	484	2 S7602	ccog protein - par
27	69	12.3	229	2 C43330	gene 7 protein - p
28	69	12.3	1111	2 B44018	laminin B2t chain
29	69	12.3	1172	2 A42587	thrombospondin 2 p

30	69	12.3	1193	2 A44018	laminin B2t chain
31	69	12.3	1557	2 T28811	hypothetical prote
32	69	12.3	1639	1 MMFEB2	laminin gamma-1 ch
33	68.5	12.2	610	2 A35046	E-selectin precurs
34	68.5	12.2	2225	2 T26063	hypothetical prote
35	68	12.1	583	2 T17326	hypothetical prote
36	68	12.1	932	2 I52527	hypothetical prote
37	68	12.1	962	2 J05571	PACE4A - mouse (fr
38	67.5	12.1	473	2 T49283	subtilisin-like pr
39	67.5	12.1	1895	2 T15881	ADAM 4 protein pre
40	67	12.0	1713	2 A55347	hypothetical prote
41	67	12.0	1766	2 A42125	adhesive ligand ep
42	66.5	11.9	435	2 I54182	trophozoite cystel
43	66.5	11.9	495	2 G82371	tumor necrosis fac
44	66.5	11.9	617	2 T49444	Fixg-related prote
45	66.5	11.9	1191	2 T13850	lustrin A related gene u-shaped prot

ALIGNMENTS

RESULT 1
A46031
Immobilization surface I-antigen precursor - Ichthyophthirius multifiliis (fragment)
C:Species: Ichthyophthirius multifiliis
C:Date: 29-Sep-1999 #sequence_revision 29-Sep-1999 #text_change 07-Dec-1999
C:Accession: A46031
R:Clark, T.G.; McGraw, R.A.; Dickerson, H.W.
Proc. Natl. Acad. Sci. U.S.A. 89, 6363-6367, 1992
A:Title: Developmental expression of surface antigen genes in the parasitic ciliate I
A:Reference number: A46031; MUID:92335298
A:Accession: A46031
A:Molecule type: mRNA; protein
A:Residues: 1-395 <CUA>
A:Cross-references: GB:M92907; NID:93628568; PIDN:AC36158.1; PID:93628569
F:2-395/Product: Immobilization surface I-antigen #status experimental <MAP>
F:156,191,245,281/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 97.1% Score 544; DB 2; Length 395;
Best Local Similarity 96.2% Pred. No. 3.2e-44;
Matches 101; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSAOGEANGNOPFAANNAARICVPCQINRVSGTNGDLTLATQSTQCTGTALDDG 60
D:|||||
Db 37 GAAOGEANGNOPFAANNAARICVPCQINRVSGTNGDLTLATQSTQCTGTALDDG 96
QY 61 VTDFDRSAACVCKRPFYNGSSPGEAPGLQVFAAGAAAGI 105
D:|||||
Db 97 VTDFDRSAACVCKRPFYNGSSPGEAPGLQVFAAGAAAGI 141

RESULT 2
C42125
trophozoite cysteine-rich surface antigen 72 - Giardia lamblia (fragment)
N:Alternate names: CRP72
C:Species: Giardia lamblia
C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 05-Dec-1998
C:Accession: C42125
R:Adam, R.D.; Yang, Y.M.; Nash, T.E.
Mol. Cell. Biol. 12, 1194-1201, 1992
A:Title: The cysteine-rich protein gene family of Giardia lamblia: loss of the CRP70
A:Reference number: A42125; MUID:92168850
A:Accession: C42125
A:Molecule type: DNA
A:Residues: 1-677 <ADA>
A:Cross-references: GB:M83934; NID:9159123
A:Experimental source: trophozoites

OY 49 TO-----CPTGALDGVTVDFDRSAQCCKPNTYNGSGPGEAP 91
 Db 1095 TPCKRPGQPLVKEKNCQFCPRGTFFON-----EEDESTCKLCAFD--HTTAAP-GATA 1144
 OY 92 GLOVEFAAGAAAG 104
 Db 1145 ESQCFSTNOCATG 1157

RESULT 7

A36385
 surface antigen serH3 - Tetrahymena thermophila
 C:Species: Tetrahymena thermophila
 C>Date: 28-Mar-1991 #sequence_revision 28-Mar-1991 #text_change 07-Dec-1999
 C:Accession: A36385
 R:Tondravil, M.M.; Willis, R.L.; Love Jr., H.D.; Bannon, G.A.
 Mol. Cell. Biol. 10, 6091-6096, 1990
 A:Title: Molecular characterization of SerH3, a Tetrahymena thermophila gene encoding a
 A:Reference number: A36385; MUID:91042547
 A:Accession: A36385
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-439 <TON>
 A:Cross-references: GB:M60425; NID:g161856; PID:g161857
 C:Genetics:
 A:Genetic code: SGC5
 C:Keywords: surface antigen

Query Match 13.4%; Score 75; DB 2; Length 439;
 Best Local Similarity 31.1%; Pred. No. 7.3;
 Matches 33; Conservative 9; Mismatches 46; Indels 18; Gaps 5;
 OY 2 SAQGEANGNPFANNAARGICVPCQINRVSGTNA---GDLATLACQSTQC---PTGT 55
 Db 22 SAGQANCTGVAAGTDCASVGVF---TVAGTGTACSWSSSLTTCYTDCTCITGT 78
 OY 56 ALDDGVTVDFDRSAQCCKPNTYNGSGPGEAPGLQVFAAGAA 101
 Db 79 V-TGITNLNDQFCTSCKSTNTYANGA-----GTACVAAASAS 115

RESULT 8

D83130
 Probable iron-sulfur protein PA4131 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Sep-2000
 C:Accession: D83130
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
 .; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A:Reference number: A82950
 A:Accession: D83130
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-573 <STO>
 A:Cross-references: GB:AE004829; GB:AE004091; NID:g9950327; PIDN:ANG07518.1; GSPDB:GN001
 C:Genetics:
 A:Experimental source: strain PA01
 A:Gene: PA4131

Query Match 13.3%; Score 74.5; DB 2; Length 573;
 Best Local Similarity 25.0%; Pred. No. 10;
 Matches 19; Conservative 12; Mismatches 18; Indels 27; Gaps 3;
 OY 2 SAQGEANGNPFANNAARGI--CVPCQINRVSGTNAAGLTLATQCTSTOCPTGTALDD 59
 Db 350 AARBSGARARRKSGSDPRAGLGDICDQ-----QCVVCPTGIDIRD 391

OY 60 GVTDFDRSAQCCK 75
 Db 392 GL-----QIACIGC 400

RESULT 9

T21560
 hypothetical protein T25C12.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T21560; T23053; T25270
 R:Harris, B.

submitted to the EMBL Data Library, August 1996
 A:Reference number: Z19441
 A:Accession: T21560

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2014 <WIL>

A:Cross-references: EMBL:Z78543; PIDN:CAB01757.1; GSPDB:GN00028; CESP:T25C12.3

A:Experimental source: clone F2966

R:White, S.

submitted to the EMBL Data Library, April 1997

A:Reference number: Z19664

A:Accession: T23053

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2014 <WIL>

A:Cross-references: EMBL:Z65566; PIDN:CA91487.1; GSPDB:GN00028; CESP:T25C12.3

A:Experimental source: clone T25C12

C:Genetics:

A:Gene: CESP:T25C12.3

A:Map position: X

A:Introns: 10/2; 91/3; 136/1; 159/2; 182/3; 203/1; 225/3; 280/1; 422/3; 466/3; 559/3;

3; 1722/1; 1779/3; 1843/3; 1908/2; 1941/2; 2007/2

Query Match 13.3%; Score 74.5; DB 2; Length 2014;
 Best Local Similarity 26.9%; Pred. No. 32;
 Matches 28; Conservative 10; Mismatches 31; Indels 35; Gaps 6;

OY 7 ANGOFPANNAARGICVPCQINRVSGTNAAGLTLATQCTSTOCPTGTALDDGVTVDF 66
 Db 674 ADDKGFYIORTTGTFC-----SGTP-----TTPYPNSCNGCVTD--P 711
 OY 67 RSAQCCKPNTYNGSGPGEAPGLQVFAAGAA 101
 Db 712 TNAATCI-CPPGFYGYQCTNIQCVNNGTTPRG--GMCVCPVGT 751

RESULT 10

T22759
 hypothetical protein F55H12.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Aug-2000
 C:Accession: T22759
 R:Dobson, R.

submitted to the EMBL Data Library, October 1996
 A:Reference number: Z19610
 A:Accession: T22759

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2824 <WIL>

A:Cross-references: EMBL:Z81091; PIDN:CAB03143.1; GSPDB:GN00019; CESP:F55H12.3

A:Experimental source: clone F55H12

A:Cross-references: GB:M67454; NID:g182409; PIDN:AAA63174.1; PID:g182410
 R:Krammer, P.H.
 Submitted to the EMBL Data Library, February 1992
 A:Reference number: S24543
 A:Accession: S24543
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-335 <KRA>
 A:Cross-references: EMBL:X63717; NID:g28741; PID:g28742
 R:Oehm, A.; Behrmann, I.; Falk, W.; Pawlita, M.; Mater, G.; Kias, C.; Li-Weber, M.; Rich, J. Biol. Chem. 267, 10709-10715, 1992
 A:Title: Purification and molecular cloning of the APO-1 cell surface antigen, a member
 A:Reference number: A38142; MUID:92268122
 A:Accession: A38142
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-134, 'O', 136-335 <OE>
 A:Experimental source: SKW6.4 cells
 A:Note: sequence extracted from NCBI backbone (NCBIP:103810)
 A:Note: in NCBI backbone the source is designated as mouse
 C:Genetics:
 A:Gene: GDB:APR1
 A:Cross-references: GDB:132671; OMIM:134637
 A:Map position: 10q24.1-10q24.1
 C:Superfamily: NGF receptor repeat homology
 C:Keywords: apoptosis; surface antigen; transmembrane protein
 F:1-16/Domain: signal sequence #status predicted <SIG>
 F:85-128/Domain: NGF receptor repeat homology <NG4>
 F:174-190/Domain: transmembrane #status predicted <TM>

Query Match 12.7%; Score 71; DB 2; Length 335;
 Best Local Similarity 28.6%; Pred. No. 14;
 Matches 22; Conservative 10; Mismatches 27; Indels 18; Gaps 4;

OY 8 GNONPFAANNAARGICVPCQINRVSGTNAGDLATLQCC--STDCPTGTALDGCVTDF 65
 DB 76 NGDEP-----DCVPCQ-----EGKEYTDKAHFSSKRCRCRLCDGHLVEEIN--C 119
 OY 66 DRSAQCVCCKRPNFYNN 82
 DB 120 TRTQNTKRCRCRKNPFNCN 136

RESULT 15
 A45664
 variant-specific surface protein VSP1267 - Giardia lamblia
 C:Species: Giardia lamblia
 C:Date: 22-Apr-1993 #sequence.revision 18-Nov-1994 #text_change 29-Jan-1999
 C:Accession: A45664
 R:Mowatt, M.R.; Agarwal, A.; Nash, T.E.
 M.Ol. Biochem. Parasitol. 49, 215-227, 1991
 A:Title: Carboxy-terminal sequence conservation among variant-specific surface proteins
 A:Reference number: A45664; MUID:92131058
 A:Accession: A45664
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-596 <MOW>
 A:Cross-references: GB:M63966; NID:g159140; PID:g159141
 A:Note: sequence extracted from NCBI backbone (NCBIN:77609, NCBIP:77610)

Query Match 12.7%; Score 71; DB 2; Length 596;
 Best Local Similarity 23.0%; Pred. No. 23;
 Matches 28; Conservative 9; Mismatches 37; Indels 48; Gaps 6;

OY 9 GNONPFAANNAARG---ICVPCQINRVSGTNAGD-----LAT 42
 DB 237 GKTHTPTDSAGGNKRYCVCSTNGIENGCECTSKESARAGTEITCTKSSNNLSP 296
 OY 43 LATCGSTOCPTGTALDGC-----VTDVFDRAQAQVCCKRPNF---YNG 83
 DB 297 LGDACLTDPCAGTYAVSGDSQSVCKPCHNTCAGCOTD--DRETS-CTACYPGYSLLYESN 353

OY 84 GS 85
 DB 354 GA 355

Search completed: March 6, 2001, 12:51:01
 Job time: 182 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:54:43 ; Search time 38.83 Seconds

(Without alignments)
87.326 Million cell updates/sec

Title: US-09-196-161D-10

Perfect score: 560

Sequence: 1 GSAOGEANGNOPFAANNAAR.....FOGEAPGLQVFAAGAAAGI 105

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	84.5	15.1	713	1	TSA4_GIALA
2	82	14.6	3635	1	LMA5_MOUSE
3	80	14.3	2369	1	LMA3_MOUSE
4	71.5	12.8	415	1	TNRC_MOUSE
5	71	12.7	335	1	FASA_HUMAN
6	71	12.7	969	1	PAC4_HUMAN
7	71	12.7	1639	1	LMG1_DROME
8	71	12.7	3110	1	LMR2_HUMAN
9	70.5	12.6	909	1	AD12_HUMAN
10	70.5	12.6	1607	1	LMG1_MOUSE
11	70	12.5	424	1	ASP_ANCCA
12	70	12.5	3084	1	LMR1_MOUSE
13	70	12.5	3712	1	LMA1_DROME
14	69	12.3	229	1	VG07_BPP22
15	69	12.3	1172	1	TSP2_MOUSE
16	69	12.3	1193	1	LMG2_HUMAN
17	69	12.3	1557	1	LM11_CAEBL
18	68.5	12.2	610	1	LEM2_HUMAN
19	67.5	12.1	1895	1	YLR3_CAEBL
20	67	12.0	1713	1	LMR3_HUMAN
21	66.5	11.9	435	1	TNRC_HUMAN
22	66	11.8	363	1	PGLR_ASPOR
23	66	11.8	1246	1	YMY2_CAEBL
24	65.5	11.7	123	1	NLRP_PINTA
25	65.5	11.7	352	1	AMBP_HUMAN
26	65.5	11.7	867	1	SSPO_BOVIN
27	65	11.6	541	1	YKCS_CAEBL
28	65	11.6	704	1	BRG3_DROME
29	65	11.6	834	1	TF1B_MOUSE
30	65	11.6	3075	1	LMR1_HUMAN
31	64.5	11.5	345	1	YJIN_ECOLI
32	64.5	11.5	484	1	LEM2_PIG
33	64.5	11.5	846	1	ITBX_DROME

34	64.5	11.5	1468	1	N153-RAT
35	64	11.4	363	1	PGLR_ASPRA
36	64	11.4	1609	1	LMG1_HUMAN
37	64	11.4	2038	1	FSH_DROME
38	64	11.4	2264	1	POLL_TBRSV
39	63.5	11.3	316	1	CH12_SOLTU
40	63.5	11.3	460	1	TUL3_MOUSE
41	63.5	11.3	485	1	LEM2_BOVIN
42	63.5	11.3	585	1	YKCS_CAEBL
43	63.5	11.3	666	1	HNF4_DROME
44	63.5	11.3	937	1	PAC4_RAT
45	63	11.2	251	1	COBM_MYCTU

ALIGNMENTS

RESULT	ID	TS4_GIALA	STANDARD	PRT	713 AA
AC	P21849				
DT	01-MAY-1991 (Rel. 18, Created)				
DT	01-MAY-1991 (Rel. 18, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	MAJOR SURFACE-LABELLED TROPHOZOITE ANTIGEN PRECURSOR.				
GN	TSA 417.				
OS	Giardia lamblia (Giardia intestinalis).				
OC	Eukaryota; Diplomonadida; Hexamitidae; Giardiae; Giardia.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-ATCC 30957 / WB.				
RC	MEDLINE-90280395; PubMed-2352929;				
RA	Gilllin F.D., Hagblom P., Harwood J., Alez S.B., Reiner D.S.,				
RA	McCaffery M., So M., Guiney D.G.,				
RT	"Isolation and expression of the gene for a major surface protein of				
RT	Giardia lamblia."				
RL	Proc. Natl. Acad. Sci. U.S.A. 87:4463-4467(1990).				
RP	[2]				
RP	SEQUENCE OF 480-620 FROM N.A.				
RC	STRAIN-AD-1:				
RC	MEDLINE-93314970; PubMed-8325510;				
RA	Ey P.L., Mayrhofer G.,				
RT	"Two genes encoding homologous 70-KDa surface proteins are present				
RT	within individual trophozoites of the binucleate protozoan parasite				
RT	Giardia intestinalis."				
RL	Gene 129:257-262(1993).				
CC	- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.				
CC	- DOMAIN: CONTAINS 29 REPEATS OF THE CXXC MOTIF.				
CC	-----				
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL: M33641; AAA02688.1; -				
DR	EMBL: M97486; AAA02581.1; -				
DR	PUR: A35502; A35502.				
KW	Signal; Antigen; Glycoprotein; Transmembrane; Repeat.				
FT	SIGNAL	1	17		
FT	CHAIN	18	713		
FT	DOMAIN	18	679		
FT	TRANSMEM	680	708		
FT	DOMAIN	709	713		
FT	CARBOHYD	289	289		
FT	CARBOHYD	676	676		
FT	VARIANT	582	582		
FT	VARIANT	606	606		
SQ	SEQUENCE	713 AA:	72510 MW:		
			9ADV1958430E5601 CRC64;		

FT	DISULFID	417	429	BY SIMILARITY.
FT	DISULFID	419	438	BY SIMILARITY.
FT	DISULFID	449	449	BY SIMILARITY.
FT	DISULFID	452	461	BY SIMILARITY.
FT	DISULFID	464	476	BY SIMILARITY.
FT	DISULFID	466	483	BY SIMILARITY.
FT	DISULFID	485	494	BY SIMILARITY.
FT	DISULFID	497	507	BY SIMILARITY.
FT	DISULFID	510	522	BY SIMILARITY.
FT	DISULFID	512	528	BY SIMILARITY.
FT	DISULFID	530	539	BY SIMILARITY.
FT	DISULFID	542	552	BY SIMILARITY.
FT	DISULFID	555	567	BY SIMILARITY.
FT	DISULFID	557	573	BY SIMILARITY.
FT	DISULFID	575	584	BY SIMILARITY.
FT	DISULFID	587	597	BY SIMILARITY.
FT	DISULFID	600	612	BY SIMILARITY.
FT	DISULFID	602	619	BY SIMILARITY.
FT	DISULFID	621	630	BY SIMILARITY.
FT	DISULFID	633	643	BY SIMILARITY.
FT	DISULFID	1360	1372	BY SIMILARITY.
FT	DISULFID	1362	1379	BY SIMILARITY.
FT	DISULFID	1381	1390	BY SIMILARITY.
FT	DISULFID	1393	1403	BY SIMILARITY.
FT	DISULFID	1450	1465	BY SIMILARITY.
FT	DISULFID	1452	1472	BY SIMILARITY.
FT	DISULFID	1474	1483	BY SIMILARITY.
FT	DISULFID	1486	1496	BY SIMILARITY.
FT	DISULFID	1499	1511	BY SIMILARITY.
FT	DISULFID	1501	1518	BY SIMILARITY.
FT	DISULFID	1520	1529	BY SIMILARITY.
FT	DISULFID	1532	1547	BY SIMILARITY.
FT	DISULFID	1782	1791	BY SIMILARITY.
FT	DISULFID	1784	1798	BY SIMILARITY.
FT	DISULFID	1801	1810	BY SIMILARITY.
FT	DISULFID	1813	1829	BY SIMILARITY.
FT	DISULFID	1832	1847	BY SIMILARITY.
FT	DISULFID	1834	1856	BY SIMILARITY.
FT	DISULFID	1858	1867	BY SIMILARITY.
FT	DISULFID	1870	1885	BY SIMILARITY.
FT	DISULFID	1888	1903	BY SIMILARITY.
FT	DISULFID	1890	1910	BY SIMILARITY.
FT	DISULFID	1913	1922	BY SIMILARITY.
FT	DISULFID	1925	1939	BY SIMILARITY.
FT	DISULFID	1989	2000	BY SIMILARITY.
FT	DISULFID	1991	2007	BY SIMILARITY.
FT	DISULFID	2009	2018	BY SIMILARITY.
FT	DISULFID	2021	2033	BY SIMILARITY.
FT	DISULFID	2036	2043	BY SIMILARITY.
FT	DISULFID	2038	2050	BY SIMILARITY.
FT	DISULFID	2052	2061	BY SIMILARITY.
FT	DISULFID	2064	2083	BY SIMILARITY.
FT	DISULFID	2086	2086	INTERCHAIN (PROBABLE).
FT	DISULFID	2089	2089	INTERCHAIN (PROBABLE).
FT	CARBOHYD	17	17	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	65	65	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	165	165	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	300	300	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	374	374	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	402	402	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	822	822	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	843	843	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	881	881	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1252	1252	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1451	1451	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1938	1938	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2115	2115	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2128	2128	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2282	2282	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2312	2312	N-LINKED (GLCNAC. . .)

Query Match 14.68; Score 82; DB 1; Length 3635;
Best Local Similarity 31.7%; Pred. No. 3.5;

Matches	26;	Conservative	7;	Mismatches	29;	Indels	20;	Gaps	4;
Oy	21	GICVPCQINFRVS-----GTNAGDLATLATQC-STQCP-----TGTAIDGGTGD	63						
Db	1796	GICVGCQINTGDCERCPRGPFVSSDPSPASPCVCPPLAVSNMFRADGCVLNHRQ	1855						
Oy	64	VEDR---SAOCVCKCPNRYN	82						
Db	1856	CLCRGYAGASCERCARPGFPGN	1877						
RESULT	3								
LMA3_MOUSE									
ID	LMA3_MOUSE	STANDARD:	PRT:	2569	AA.				
AC	061788;	061788;	061966;						
DT	01-NOV-1997	(Rel. 35, Created)							
DT	01-NOV-1997	(Rel. 35, Last sequence update)							
DT	15-JUL-1999	(Rel. 38, Last annotation update)							
DE	LAMININ ALPHA-3 CHAIN	PRECURSOR (FRAGMENT).							
GN	LMA3.								
OS	Mus musculus (Mouse).								
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RC	SPRAIN-BALB/C; TISSUE=LUNG;								
RX	MEDLINE=95394948; PubMed=7665604;								
RA	Galliano M.-F., Aberdam D., Aguzzi A., Ortonne J.-P., Meneguzzi G.;								
RT	"Cloning and complete primary structure of the mouse laminin alpha 3								
RT	chain. Distinct expression pattern of the laminin alpha 3A and alpha								
RT	3B chain isoforms.";								
RL	J. Biol. Chem. 270:21820-21826(1995).								
RN	[2]								
RP	REVISIONS.								
RA	Aberdam D.;								
RL	Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.								
RN	[3]								
RP	SEQUENCE OF 1052-1770 FROM N.A.								
RC	TISSUE=LUNG;								
RX	MEDLINE=94281750; PubMed=8012114;								
RA	Aberdam D., Galliano M.-F., Mattei M.-G., Pisani-Spadafora A.,								
RT	Ortonne J.-P., Meneguzzi G.;								
RT	"Assignment of mouse nlecin genes to chromosomes 1 and 18.";								
RN	Mamm. Genome 5:229-233(1994).								
RP	[4]								
RC	SEQUENCE OF 1052-1770 FROM N.A.								
RC	TISSUE=LUNG;								
RX	MEDLINE=94363405; PubMed=8081888;								
RA	Aberdam D., Aguzzi A., Baudoin C., Galliano M.-F., Ortonne J.-P.,								
RA	Meneguzzi G.;								
RT	"Developmental expression of nlecin adhesion protein (laminin-5)								
RT	subunits suggests multiple morphogenic roles.";								
RL	Cell Adhes. Commun. 2:115-129(1994).								
CC	-1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR. LAMININ								
CC	IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF								
CC	CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING								
CC	WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.								
CC	-1- FUNCTION: LAMININ-5 IS THOUGHT TO BE INVOLVED IN (1) CELL ADHESION								
CC	VIA INTEGRIN ALPHA-3/BETA-1 IN FOCAL ADHESION AND INTEGRIN ALPHA-								
CC	6/BETA-4 IN HEMIDESMOSOMES, (2) SIGNAL TRANSDUCTION VIA TYROSINE								
CC	PHOSPHORYLATION OF P125-FAK AND P80, (3) DIFFERENTIATION OF								
CC	KERATINOCYTES (BY SIMILARITY).								
CC	-1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE								
CC	DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND								
CC	TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE								
CC	COMPRISED ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.								
CC	THE ALPHA-3 CHAIN IS A SUBUNIT OF LAMININ-5 (EPITIGRIN/KALININ/								
CC	NICEIN), AND POSSIBLY ALSO A COMPONENT OF LAMININ-6 (K-LAMININ)								
CC	AND LAMININ-7 (KS-LAMININ).								
CC	-1- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT								
CC	MEMBRANES (MAJOR COMPONENT).								
CC	-1- ALTERNATIVE PRODUCTS: THE TWO ISOFORMS A AND B, WHICH DIFFER								
CC	IN THEIR N-TERMINAL ARE DERIVED BY ALTERNATIVE SPLICING OF THE								

CC SAME GENE. THE SEQUENCE SHOWN HERE IS THAT OF THE LARGER ISOFORM B.

CC - TISSUE SPECIFICITY: BASAL MEMBRANE OF THE UPPER ALIMENTARY TRACT AND URINARY AND NASAL EPITHELIA, SALIVARY GLANDS AND TEETH (BOTH VARIANTS). ISOFORM A IS PREDOMINANTLY EXPRESSED IN SKIN, HAIR FOLLICLES AND DEVELOPING NEURONS OF THE TRIGEMINAL GANGLION.

CC ISOFORM B WAS FOUND IN BRONCHI, ALVEOLI, STOMACH, INTESTINAL CRYPTS, WHISKER PADS, CNS, TELENCEPHALIC NEUROECTODERM, THALAMUS, RATHKE'S POUCH, AND PERIVENTRICULAR SUBEPIDYMAL GERMINAL LAYER.

CC - DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.

CC - DOMAIN: DOMAINS IV AND G ARE GLOBULAR.

CC - SIMILARITY: CONTAINS 6.5 LAMININ EGF-LIKE DOMAINS.

CC - SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV (DOMAIN IV' IS NOT SIMILAR TO LAMININ DOMAIN IV).

CC - SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb.ch/announce/or_send_an_email_to_license@isb-sib.ch).

CC -----

CC EMBL: X84014; CAA58837.1; -

CC EMBL: X84013; CAA58836.1; -

CC EMBL: L20478; AAA68091.1; -

CC HSSP: P02468; 1TLE.

CC MGD: MGI:99909; LAMA3.

CC INTERPRO: IPR000034; -

CC INTERPRO: IPR000561; -

CC INTERPRO: IPR001791; -

CC INTERPRO: IPR002049; -

CC PFAM: PF000052; laminin_B; 1.

CC PFAM: PF000053; laminin_EGF; 4.

CC PFAM: PF000054; laminin_G; 3.

CC PROSITE: PS00022; EGF_1; 4.

CC PROSITE: PS00186; EGF_2; 1.

CC PROSITE: PS01248; LAMININ_TYPE_EGF; 4.

CC Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil; Laminin EGF-like domain; Cell adhesion; Repeat; Signal;

KW Alternative splicing.

KW NON_TER 1

FT SIGNAL 1 28

FT CHAIN 29 2569

FT DOMAIN 29 498

FT DOMAIN 499 700

FT DOMAIN 546 700

FT DOMAIN 546 589

FT DOMAIN 590 639

FT DOMAIN 640 690

FT DOMAIN 691 700

FT DOMAIN 701 889

FT DOMAIN 890 1057

FT DOMAIN 890 922

FT DOMAIN 923 969

FT DOMAIN 970 1022

FT DOMAIN 1023 1057

FT DOMAIN 1058 1648

FT DOMAIN 1649 2569

FT DOMAIN 1649 1825

FT DOMAIN 1826 1994

FT DOMAIN 1995 2209

FT DOMAIN 2210 2385

FT DOMAIN 2386 2569

FT DOMAIN 1090 1251

FT DOMAIN 1251 1296

FT DOMAIN 1327 1404

FT DOMAIN 1450 1477

FT DOMAIN 1557 1622

FT SITE 1513 1515

POTENTIAL.

LAMININ ALPHA-3 CHAIN.

DOMAIN IV'.

DOMAIN III B.

3.5 X LAMININ EGF-LIKE REPEATS.

LAMININ EGF-LIKE 1.

LAMININ EGF-LIKE 2.

LAMININ EGF-LIKE 3.

LAMININ EGF-LIKE 4 (N-TERMINAL).

LAMININ DOMAIN IV (DOMAIN IV A).

3 X LAMININ EGF-LIKE REPEATS (DOMAIN III A).

LAMININ EGF-LIKE 4 (C-TERMINAL).

LAMININ EGF-LIKE 5.

LAMININ EGF-LIKE 6.

LAMININ EGF-LIKE 7 (INCOMPLETE).

DOMAIN II AND I (HEPTAR REPEATS).

5 X LAMININ G-LIKE REPEATS (DOMAIN G).

LAMININ G-LIKE 1.

LAMININ G-LIKE 2.

LAMININ G-LIKE 3.

LAMININ G-LIKE 4.

LAMININ G-LIKE 5.

COILED COIL (POTENTIAL).

COILED COIL (POTENTIAL).

COILED COIL (POTENTIAL).

COILED COIL (POTENTIAL).

CELL ATTACHMENT SITE (POTENTIAL).

FT DISULFID 546 553 BY SIMILARITY.

FT DISULFID 548 560 BY SIMILARITY.

FT DISULFID 562 571 BY SIMILARITY.

FT DISULFID 574 587 BY SIMILARITY.

FT DISULFID 590 605 BY SIMILARITY.

FT DISULFID 592 612 BY SIMILARITY.

FT DISULFID 614 623 BY SIMILARITY.

FT DISULFID 626 637 BY SIMILARITY.

FT DISULFID 640 652 BY SIMILARITY.

FT DISULFID 642 659 BY SIMILARITY.

FT DISULFID 661 670 BY SIMILARITY.

FT DISULFID 673 688 BY SIMILARITY.

FT DISULFID 923 932 BY SIMILARITY.

FT DISULFID 925 939 BY SIMILARITY.

FT DISULFID 942 951 BY SIMILARITY.

FT DISULFID 954 967 BY SIMILARITY.

FT DISULFID 970 982 BY SIMILARITY.

FT DISULFID 972 991 BY SIMILARITY.

FT DISULFID 993 1002 BY SIMILARITY.

FT DISULFID 1005 1020 BY SIMILARITY.

FT DISULFID 1058 1058 BY SIMILARITY.

FT DISULFID 1061 1061 INTERCHAIN (PROBABLE).

FT CARBOHYD 591 591 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 912 912 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1398 1398 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1500 1500 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1571 1571 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1600 1600 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1737 1737 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1819 1819 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1986 1986 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2333 2333 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2509 2509 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT VARSPIC 1 842 MISSING (IN ISOFORM A).

FT VARSPIC 843 901 MIVLPRLERHIGSLGTEORLIGVLEPASDPSGGR

FT AHLEVMCACPDYTGDSG -> MIPAYRMSMSTGMLITG

FT AALGQCGVGSSEQRFVAFIQRPSQNHQAQSMELRPS (IN ISOFORM A).

FT SQ SEQUENCE 2569 AA; 282159 MW; 40D4306BEF340DBC CRC64;

Query Match 14.3%; Score 80; DB 1; Length 2569;

Best Local Similarity 29.4%; Pred. No. 4;

Matches 25; Conservative 7; Mismatches 35; Indels 18; Gaps 4;

QY 21 GICVPCOINRGV---SGTNAGDLATLTCSTGCP-----TGTAADDGVTVDFR-- 67

DB 937 GICINCQHTAGCEHCRCQACHYGNALHSGRCVCPCHTNSFAFGCAVDGAVNCAKCKPG 996

QY 68 -SAAQCVKCKRNFYFN---GGSPO 87

DB 997 YTGQCERCACGVCYGNPKFGSGSQ 1021

RESULT 4

TNRC_MOUSE 1

ID TNRC_MOUSE STANDARD; PRT; 415 AA.

AC P50284;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE LYMPHOTOXIN-BETA RECEPTOR PRECURSOR.

GN LYBR OR TNFR.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

RP [1]

RC SEQUENCE FROM N.A.

RC STRAIN-CVB: TISSUE=LUNG;

RX MEDLINE=96072804; PubMed=7594541;

RA Force W.R., Walter B.N., Hession C., Tizard R., Kozak C.A.,

RA Browning J.L., Ware C.F.;

RA "Mouse lymphotoxin-beta receptor. Molecular genetics, ligand binding,

RT and expression.";
 RL J. Immunol. 155:5280-5288(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9616385; Pubmed=8586432;
 RA Nakamura T., Tashiro K., Nazarea M., Nakano T., Sasayama S.,
 RA Honjo T.;
 RT "The murine lymphotoxin-beta receptor cDNA: isolation by the signal
 sequence trap and chromosomal mapping.";
 RL Genomics 30:312-319(1995).
 CC -1- FUNCTION: RECEPTOR FOR THE LYMPHOTOXIN-BETA. POSSIBLE FUNCTION IN
 CC IMMUNE DEVELOPMENT.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- SIMILARITY: CONTAINS A LA-NGFR/TNFR-TYPE CYSTEINE-RICH REGION.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U29173; AAA68964.1; -;
 DR EMBL: L38423; AAB00846.1; -;
 DR EMBL: U30798; AAA81334.1; -;
 DR HSSP: P25942; ICDF.
 DR MGD: MGI:104875; LTRB.
 DR INTERPRO: IPR001368; -;
 DR PRAM: PF00020; TNFR_C6; 3.
 DR PROSITE: PS00652; TNFR_NGFR_1; 2.
 DR PROSITE: PS50050; TNFR_NGFR_2; 3.
 KM Receptor; Transmembrane; Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 30
 FT CHAIN 31 415
 FT DOMAIN 31 223
 FT TRANSMEM 224 244
 FT DOMAIN 245 415
 FT DOMAIN 42 213
 FT REPEAT 42 81
 FT REPEAT 82 124
 FT REPEAT 125 170
 FT REPEAT 171 213
 FT REPEAT 43 58
 FT DISULEID 59 72
 FT DISULEID 62 80
 FT DISULEID 83 98
 FT DISULEID 101 116
 FT DISULEID 104 124
 FT DISULEID 126 132
 FT DISULEID 139 150
 FT DISULEID 142 169
 FT DISULEID 172 187
 FT CARBOHYD 40 179
 FT CARBOHYD 179 179
 SQ SEQUENCE 415 AA; 44956 MW; 298326a566aef661 CRC64;
 Query Match 12.88; Score 71.5; DB 1; Length 415;
 Best Local Similarity 26.48; Pired. No. 4.9;
 Matches 23; Conservative 7; Mismatches 30; Indels 27; Gaps 3;
 QY 23 CVCQINRVSGTINAGDLATLQCTGPTALDGVTVDFVDSAAQCVCKPKPNFYNN 82
 Db 139 CVCHEERL-----VLQCPTEAE--VDEIMDTIVNCVPCPKPGHFN 179
 QY 83 GGSFG-----EAPGLQVPAAGAA 101
 Db 180 TSSPRACOPHTRCEIQLGLEAAGT 206
 RESULT 5
 FASA_HUMAN

ID FASA_HUMAN STANDARD; PRT; 335 AA.
 AC P25445;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE FASL RECEPTOR PRECURSOR (APOPTOSIS-MEDIATING SURFACE ANTIGEN FAS)
 DE (APO-1 ANTIGEN) (CD95 ANTIGEN).
 GN TNFRSF6 OR APT1 OR FAS OR FASL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91309137; Pubmed=1713127;
 RA Itoh N., Yonehara S., Ishii A., Yonehara S.I.,
 RA Sameshima M., Hase A., Seto Y., Nagata S.;
 RT "The polypeptide encoded by the cDNA for human cell surface antigen
 RT Fas can mediate apoptosis.";
 RL Cell 66:233-243(1991).
 RN [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 226-240; 269-291 AND 321-335.
 RX MEDLINE=92268123; Pubmed=1375228;
 RA Oehm A., Behrmann I., Falk W., Pawlita M., Walter G., Klas C.,
 RA Li-Weber M., Richards S., Dhein J., Trauth B.C., Ponsstingl H.,
 RA Krammer P.H.;
 RT "Purification and molecular cloning of the Apo-1 cell surface
 RT antigen, a member of the tumor necrosis factor/nerve growth factor
 RT receptor superfamily. Sequence identity with the Fas antigen.";
 RL J. Biol. Chem. 267:10709-10715(1992).
 RN [3]
 RP STRUCTURE BY NMR OF 218-335.
 RX MEDLINE=97122332; Pubmed=8967952;
 RA Huang B., Eberstadt M., Olejniczak E.T., Meadows R.P., Fesik S.W.;
 RT "NMR structure and mutagenesis of the Fas (Apo-1/CD95) death domain.";
 RL Nature 384:638-641(1996).
 CC -1- FUNCTION: RECEPTOR FOR A CYTOKINE LIGAND KNOWN AS FASL. THE
 CC ADAPTOR MOLECULE FADD RECRUITS CASPASE-8 TO THE ACTIVATED
 CC RECEPTOR. THE RESULTING AGGREGATE CALLED THE DEATH-INDUCING
 CC SIGNALING COMPLEX (DISC) PERFORMS CASPASE-8 PROTEOLYTIC
 CC ACTIVATION. ACTIVE CASPASE-8 INITIATES THE SUBSEQUENT CASCADE OF
 CC CASPASES (ASPARTATE-SPECIFIC CYSTEINE PROTEASES) MEDIATING
 CC APOPTOSIS. FAS-MEDIATED APOPTOSIS MAY HAVE A ROLE IN THE
 CC INDUCTION OF PERIPHERAL TOLERANCE, IN THE ANTIGEN-STIMULATED
 CC SUICIDE OF MATURE T-CELLS, OR BOTH.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- DOMAIN: CONTAINS A DEATH DOMAIN INVOLVED IN THE BINDING OF FADD,
 CC AND MAYBE TO OTHER CYTOSOLIC ADAPTOR PROTEINS.
 CC -1- SIMILARITY: CONTAINS A LA-NGFR/TNFR-TYPE CYSTEINE-RICH REGION.
 CC -1- SIMILARITY: CONTAINS 1 DEATH DOMAIN.
 CC -1- DATABASE: NAME=PROV; NOTE=CD guide CD95 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd95.htm"
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M67454; AAA63174.1; -;
 DR EMBL: X63717; CAA45250.1; -;
 DR PIR: A40036; A40036.
 DR PIR: S24543; S24543.
 DR PDB: 1DDF; 12-NOV-97.
 DR MIM: 134637; -;
 DR INTERPRO: IPR000488; -;
 DR INTERPRO: IPR001368; -;
 DR PRAM: PF00020; TNFR_C6; 2.
 DR PRAM: PF00531; death; 1.
 DR PROSITE: PS00652; TNFR_NGFR_1; 2.
 DR PROSITE: PS50050; TNFR_NGFR_2; 2.
 DR PROSITE: PS50017; DEATH_DOMAIN; 1.

```

KW Apoptosis/Receptor; Glycoprotein; Transmembrane; Repeat; Signal;
3d-structure.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 335 FAST RECEPTOR.
FT DOMAIN 17 173 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 174 190 POTENTIAL.
FT DOMAIN 191 335 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 47 166 3 x TNFR-CYS.
FT REPEAT 47 83 TNFR-CYS 1.
FT REPEAT 84 127 TNFR-CYS 2.
FT REPEAT 128 166 TNFR-CYS 3.
FT DOMAIN 230 314 DEATH DOMAIN.
FT CARBOHYD 118 118 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 136 136 N-LINKED (GLCNAc. . .) (POTENTIAL).
SQ SEQUENCE 335 AA: 37732 MW: 0139942535111410 CRC64.

Query Match 12.7%; Score 71; DB 1; Length 335;
Best Local Similarity 28.6%; Pred. No. 4.5;
Matches 22; Conservative 10; Mismatches 27; Indels 18; Gaps 4;

OY 8 NGNPFANNAARGICVPCQINRVSGGTNAGDLATLATQC--STGCPGTALDDGVTVDF 65
DB 76 NGDEP-----DCVPE-----EGKEVTDKAHFSSKRCRCRLCDGSHGLEVEIN--C 119
OY 66 DRSAQCCKCRKPEFYNN 82
DB 120 TRTONTKCRCKRPNFCN 136

RESULT 6
PAC4_HUMAN STANDARD: PRT: 969 AA.
AC P29122:
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SUBTILISIN-LIKE PROTEASE PAC4 PRECURSOR (EC 3.4.21.-).
GN PAC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP RP SEQUENCE FROM N.A.
RX MEDLINE=92075167; Pubmed=1741956;
RX Kiefer M.C., Tucker J.E., Joh R., Landsberg K.E., Saltman D.,
RX Barr P.J.;
RT "Identification of a second human subtilisin-like protease gene in
RT the fes/fes region of chromosome 15.";
RL DNA Cell Biol. 10:757-769(1991).
RN [2]
RP RP SEQUENCE FROM N.A. (PAC4B).
RX MEDLINE=98021085; Pubmed=9378725;
RX Tsuji A., Hine C., Tamai Y., Yonemoto K., Mori K., Yoshida S.,
RX Bando M., Sakai F., Mori K., Akamatsu T., Matsuda Y.;
RT "Genomic organization and alternative splicing of human PAC4 (SPC4),
RT kexin-like processing endoprotease.";
RL J. Biochem. 122:438-452(1997).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- ALTERNATIVE PRODUCTS: TWO VARIANTS SEEM TO BE PRODUCED BY
CC ALTERNATIVE SPLICING OF A SINGLE GENE. THEY DIFFER AFTER
CC LYS-471.
CC -1- TISSUE SPECIFICITY: WIDESPREAD, WITH COMPARATIVELY HIGHER LEVELS
CC IN THE LIVER. PAC4.1 WAS ONLY FOUND IN THE LIVER.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY. HIGH SIMILARITY WITH OTHER FURIN-LIKE ENZYMES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

```

CC      or send an email to license@lsb-sib.ch.
CC
DR      EMBL, M80482; AAs59998.1; -.
DR      EMBL, AB001914; BAA21620.1; -.
DR      EMBL, AB001898; BAA21620.1; JOINED.
DR      EMBL, AB001900; BAA21620.1; JOINED.
DR      EMBL, AB001901; BAA21620.1; JOINED.
DR      EMBL, AB001902; BAA21620.1; JOINED.
DR      EMBL, AB001903; BAA21620.1; JOINED.
DR      EMBL, AB001904; BAA21620.1; JOINED.
DR      EMBL, AB001905; BAA21620.1; JOINED.
DR      PIR, A39490; A39490.
DR      HSSP, O99405; 1MPF.
DR      MEROPS; S08.075; -.
DR      MIM; 167405; -.
DR      INTERPRO; IPR000209; -.
DR      INTERPRO; IPR002884; -.
DR      PFAM; PF01483; P.1.
DR      PFAM; PF00082; peptidase_s8; 1.
DR      PRINTS; PRO0723; SUBTILISIN.
DR      PROSITE; PS00136; SUBTILASE_ASP; 1.
DR      PROSITE; PS00137; SUBTILASE_HIS; 1.
DR      PROSITE; PS00138; SUBTILASE_SER; 1.
KW      Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
KW      Alternative splicing; Repeat.
FT      SIGNAL          1      63      POTENTIAL.
FT      PROPEL          64      149      POTENTIAL.
FT      CHAIN           150      969      SUBTILISIN-LIKE PROTEASE PACE4.
FT      DOMAIN          695      969      CYS-RICH REGION.
FT      ACT_SITE        205      205      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT      ACT_SITE        246      246      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT      ACT_SITE        420      420      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT      SITE            553      555      CELL ATTACHMENT SITE (POTENTIAL).
FT      DOMAIN          686      941      5 X TANDEM REPEATS, CYS-RICH.
FT      REPEAT          686      739      1.
FT      REPEAT          740      790      2.
FT      REPEAT          791      838      3.
FT      REPEAT          839      887      4.
FT      REPEAT          888      941      5.
FT      CARBOHYD        259      259      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD        914      914      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD        932      932      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      VARSPLIC        471      471      K -> KGAAVAFMWITGPMNV (IN ISOFORM
FT                                     PACE4.1).
FT      VARSPLIC        472      969      MISSING (IN ISOFORM PACE4.1).
FT      SEQUENCE        969 AA; 106419 MW; A3599CC278D09B05 CRC64;
SQ
Query Match          12.7%; Score 71; DB 1; Length 969;
Best Local Similarity 25.0%; Pred. No. 12;
Matches 23; Conservative 13; Mismatches 34; Indels 22; Gaps 3;
QY      12 PEAANNAARGICVPPQINNVGSGSTNA-----GDLATLATQCSNOCPTGTALD-- 58
      | : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
DbB      664 PGSANIILQTSVHPHCPCGDKCGDPAAADCLNCVHFSLSLSVKT-SKKCVSYVCLGTFGDTA 742
QY      59 -----DGVTVDFEDRSAACQCVKCKPNEYFN 82
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      743 ARRCRCHKGCETCSSRAATQCLSCRGRGYHH 774
RESULT 7
LMG1_DROME          STANDARD:          PRT; 1639 AA.
AC      P15215; Q24373; Q9VT18;
DT      01-APR-1980 (Rel. 14, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      01-OCT-2000 (Rel. 40, Last annotation update)
DE      LAMININ GAMMA-1 CHAIN PRECURSOR (LAMININ B2 CHAIN).
GN      LAMB2 OR LAMC1 OR LAMG1.
OS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC      Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

```


OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CANTON-S; AND OREGON-R;
 RX MEDLINE=91299161; PubMed=1840513;
 RA Chi H.-C., Juminaga D., Wang S.-Y., Hui C.-F.;
 RT "Structure of the Drosophila gene for the laminin B2 chain.";
 RL DNA Cell Biol. 10:451-466(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OREGON-R;
 RX MEDLINE=89109164; PubMed=2912972;
 RA Chi H.-C., Hui C.-F.;
 RT "Primary structure of the Drosophila laminin B2 chain and comparison
 with human, mouse, and Drosophila laminin B1 and B2 chains.";
 RL J. Biol. Chem. 264:1543-1550(1989).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90037237; PubMed=2808533;
 RA Montell D.J., Goodman C.S.;
 RT "Drosophila laminin: sequence of B2 subunit and expression of all
 three subunits during embryogenesis.";
 RL J. Cell Biol. 109:2441-2453(1989).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 AMannates P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman J.P., Bhandari D., Boltsakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,
 RA Burrus K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Crowley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Goad C.A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heilmann T.J., Hernandez J.R., Houck C.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iqeyam C.,
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclab J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Palmett K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Sine B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 RN [5]
 RP SEQUENCE OF 344-1639 FROM N.A.
 RC STRAIN=OREGON-R;
 RX MEDLINE=88303364; PubMed=3405777;
 RA Chi H.-C., Hui C.-F.;
 RT "cDNA and amino acid sequences of Drosophila laminin B2 chain";
 RL Nucleic Acids Res. 16:7205-7205(1988).
 CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ

CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
 CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
 CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
 CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
 CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
 CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
 CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
 CC -1- SUBCELLULAR LOCATION: BASEMENT MEMBRANES (MAJOR COMPONENT).
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -1- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
 CC -1- SIMILARITY: CONTAINS 11 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
 CC -1- SIMILARITY: CONTAINS 1 LAMININ EGF-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M58417; AAA28665.1; -;
 CC EMBL: M25063; AAA28664.1; -;
 CC EMBL: AE003551; AAF50238.1; -;
 CC EMBL: X07806; CAA30665.1; -;
 CC PIR: A31483; MMEFB2.
 CC HSSP: P02468; 1TLE.
 CC EMBASE: FBgn0002528; Lamb2.
 CC INTERPRO: IPR000034; -;
 CC INTERPRO: IPR000561; -;
 CC INTERPRO: IPR001886; -;
 CC INTERPRO: IPR002049; -;
 CC PRAM: PR000552; laminin_B; 1.
 CC PRAM: PF00053; laminin_EGF; 10.
 CC PRAM: PF00055; laminin_Nterm; 1.
 CC PROSITE: PS00022; EGF_1; 8.
 CC PROSITE: PS01186; EGF_2; 1.
 CC PROSITE: PS01248; LAMININ-TYPE EGF; 11.
 CC Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 CC Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 CC SIGNAL 1 33
 CC CHAIN 34 1639
 CC DOMAIN 34 298
 CC DOMAIN 299 523
 CC FT 299 358
 CC FT 359 413
 CC FT 414 460
 CC FT 461 513
 CC FT 514 523
 CC FT 524 709
 CC FT 710 1049
 CC DOMAIN 710 743
 CC DOMAIN 744 792
 CC DOMAIN 793 846
 CC DOMAIN 847 901
 CC DOMAIN 902 955
 CC DOMAIN 956 1003
 CC DOMAIN 1004 1049
 CC DOMAIN 1050 1609
 CC DOMAIN 1087 1109
 CC DOMAIN 1144 1247
 CC DOMAIN 1306 1627
 CC DOMAIN 299 308
 CC DISULFID 301 322
 CC DISULFID 324 333
 CC DISULFID 336 356
 CC DISULFID 359 368
 CC DISULFID 361 384
 CC DISULFID 387 396
 CC LAMININ EGF-LIKE 5 (C-TERMINAL).
 CC LAMININ EGF-LIKE 6.
 CC LAMININ EGF-LIKE 7.
 CC LAMININ EGF-LIKE 8.
 CC LAMININ EGF-LIKE 9.
 CC LAMININ EGF-LIKE 10.
 CC LAMININ EGF-LIKE 11.
 CC DOMAIN II AND I.
 CC COILED COIL (POTENTIAL).
 CC COILED COIL (POTENTIAL).
 CC COILED COIL (POTENTIAL).
 CC BY SIMILARITY.
 CC BY SIMILARITY.
 CC BY SIMILARITY.
 CC BY SIMILARITY.
 CC BY SIMILARITY.
 CC BY SIMILARITY.

```

FT DISULFID 399 411 BY SIMILARITY.
FT DISULFID 414 426 BY SIMILARITY.
FT DISULFID 416 432 BY SIMILARITY.
FT DISULFID 434 443 BY SIMILARITY.
FT DISULFID 446 458 BY SIMILARITY.
FT DISULFID 461 475 BY SIMILARITY.
FT DISULFID 463 482 BY SIMILARITY.
FT DISULFID 484 493 BY SIMILARITY.
FT DISULFID 496 511 BY SIMILARITY.
FT DISULFID 744 753 BY SIMILARITY.
FT DISULFID 746 760 BY SIMILARITY.
FT DISULFID 762 771 BY SIMILARITY.
FT DISULFID 774 790 BY SIMILARITY.
FT DISULFID 793 801 BY SIMILARITY.
FT DISULFID 795 811 BY SIMILARITY.
FT DISULFID 814 823 BY SIMILARITY.
FT DISULFID 826 844 BY SIMILARITY.
FT DISULFID 847 861 BY SIMILARITY.
FT DISULFID 849 868 BY SIMILARITY.
FT DISULFID 871 880 BY SIMILARITY.
FT DISULFID 883 899 BY SIMILARITY.
FT DISULFID 902 919 BY SIMILARITY.
FT DISULFID 904 926 BY SIMILARITY.
FT DISULFID 928 937 BY SIMILARITY.
FT DISULFID 940 953 BY SIMILARITY.
FT DISULFID 956 968 BY SIMILARITY.
FT DISULFID 958 975 BY SIMILARITY.
FT DISULFID 977 986 BY SIMILARITY.
FT DISULFID 989 1001 BY SIMILARITY.
FT DISULFID 1004 1016 BY SIMILARITY.
FT DISULFID 1006 1022 BY SIMILARITY.
FT DISULFID 1024 1033 BY SIMILARITY.
FT DISULFID 1036 1047 BY SIMILARITY.
FT DISULFID 1050 1050 INTERCHAIN (PROBABLE).
FT DISULFID 1053 1053 INTERCHAIN (PROBABLE).
FT DISULFID 1631 1631 INTERCHAIN (PROBABLE).
FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 376 376 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 669 669 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 862 862 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1070 1070 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1156 1156 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1394 1394 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

Query Match 12.7% Score 71: DB 1: Length 1639;

Best Local Similarity 29.9%; Pred. No. 20; Matches 32; Conservative 8; Mismatches 41; Indels 26; Gaps 7;

```

QY 6 EANGN-OPFANNAR--GICVPCQINRVG-----SGTNAGDLATLATQCS--P 52
      ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 848 DGNQNDVNPVAVGNCRFTGELKCIHNTAGEHCQCLSGHGDPLALPHGRCD-RCSCYE 906
      ||| | | | | | | | | | | | | | | | | | | | | | | | |
QY 53 TGTALDDGVTVDVDRSAQCYKCKPNE-----YNGGSPQG 88
      ||| | | | | | | | | | | | | | | | | | | | | | | |
DB 907 AGTEDEDSITRCDQVYGC-OCKPNVIGRDCGRCOPGYFNIRSGNG 952
      ||| | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 8

LMA2_HUMAN STANDARD: PRT; 3110 AA.

```

AC P24043; Q14736;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE LAMININ ALPHA-2 CHAIN PRECURSOR (LAMININ M CHAIN) (MEROSIN HEAVY
  CHAIN)
GN LAMA2 OR LAMM.
OS Homo sapiens (Human).
OC Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi:
  Mammalia: Eutheria: Primates: Catarrhini: Homiinae: Homo.
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC TISSUE=PLACENTA;
RX MEDLINE=94124633; PubMed=8294519;
RA Voliueano R., Nissinen M., Sainio K., Byers M., Eddy R.,
  Hirvonen H., Shows T.B., Sariola H., Engvall E., Tryggvason K.:
  "Human laminin M chain (merosin): complete primary structure,
  RT chromosomal assignment, and expression of the M and A chain in human
  RT fetal tissues.";
  RT J. Cell Biol. 124:381-394(1994).
  RL [2]
  RP SEQUENCE OF 1981-3110 FROM N.A., AND PARTIAL SEQUENCE.
  RC TISSUE=PLACENTA;
  RX MEDLINE=90238994; PubMed=2185464;
  RA Enrig K., Leivo I., Argaves W.S., Ruoslahti E., Engvall E.;
  RT "Merosin, a tissue-specific basement membrane protein, is a
  RT laminin-like protein.";
  RL Proc. Natl. Acad. Sci. U.S.A. 87:3264-3268(1990).
  RL [3]
  RP VARIANTS GLN-545; HIS-619; LEU-919; HIS-2586 AND LYS-2614.
  RA Panicker S.G., Mendell J.T., Chen L., Feng B., Sahenk Z.,
  RA Marzluft G.A., Amato A.A., Mendell J.R.;
  RT "Novel single base polymorphisms and rare sequence variants in
  RT the laminin 2-chain coding region detected by RFLP/SSCP analysis.";
  RL Hum. Mutat. 13:174-174(1999).
  RL [4]
  RP ERRATUM.
  RA Panicker S.G., Mendell J.T., Chen L., Feng B., Sahenk Z.,
  RA Marzluft G.A., Amato A.A., Mendell J.R.;
  RL Hum. Mutat. 13:340-340(1999).
  RL [5]
  CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
  CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
  CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
  CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
  CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
  CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA) WHICH ARE BOUND
  CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
  CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
  CC THE ALPHA-2 CHAIN IS A SUBUNIT OF LAMININ-2 (MEROSIN) AND LAMININ-
  CC 4 (S-MEROSIN).
  CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
  CC MEMBRANES (MAJOR COMPONENT).
  CC -1- TISSUE SPECIFICITY: PLACENTA, STRIPED MUSCLE, PERIPHERAL NERVE,
  CC CARDIAC MUSCLE, PANCREAS, LUNG, SPLEEN, KIDNEY, ADRENAL GLAND,
  CC SKIN, TESTIS, MENINGES, CHOROID PLEXUS, AND SOME OTHER REGIONS OF
  CC THE BRAIN; NOT IN LIVER, THYMUS AND BONE.
  CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
  CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
  CC -1- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
  CC -1- DISEASE: DEFECTS IN LAMA2 ARE THE CAUSE OF MEROSIN-DEFICIENT
  CC CONGENITAL MUSCULAR DYSTROPHY (MCDM).
  CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
  CC -1- SIMILARITY: CONTAINS 17 LAMININ N-TERMINAL EGF-LIKE DOMAINS.
  CC -1- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
  CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
  CC -----
  CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
  CC between the Swiss Institute of Bioinformatics and the EMBL collaboration -
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way
  CC modified and this statement is not removed. Usage by and for commercial
  CC entities requires a license agreement (See http://www.isb.ch/announce/
  CC or send an email to license@isb-sib.ch).
  CC -----
  DR EMBL: Z26653; CAA81394.1; -
  DR EMBL: M59832; AAA63215.1; -
  DR PIR: A35899; MAMHMH.
  DR HSP: P02468; IKLO.
  DR MIM: 156225; -
  DR INTERPRO: IPR000034; -
  DR INTERPRO: IPR000561; -
  DR INTERPRO: IPR001791; -
  DR INTERPRO: IPR001886; -
  DR INTERPRO: IPR002049; -
  DR PFM: PF00052; Laminin_B; 2.

```


RA Wewer U.M.;
 RT "A novel, secreted form of human ADAM 12 (meltrin alpha) provokes
 myogenesis in vivo."
 RN J. Biol. Chem. 273:157-166(1998).
 RP CHARACTERIZATION.
 RX MEDLINE=98307939; Pubmed=9642263;
 RA Loechel F., Gilpin B.J., Engvall E., Albrechtsen R., Wewer U.M.;
 RT "Human ADAM 12 (meltrin alpha) is an active metalloproteinase."
 RN J. Biol. Chem. 273:16993-16997(1998).
 CC -1- FUNCTION: MAY BE INVOLVED IN MYOBLAST FUSION AS WELL AS
 MACROPHAGE-DERIVED GIANT CELLS (MGC) AND OSTEOCLAST FORMATION FROM
 MONONUCLEAR PRECURSORS.
 CC -1- COFACTOR: BINDS ONE ZINC ION.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (ISOFORM 12L). A
 SECRETED FORM (ISOFORM 12S) IS PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: ISOFORM 12L (SHOWN HERE), AND
 ISOFORM 12S, ARE PRODUCED BY ALTERNATIVE SPLICING. ISOFORM 12S IS
 A SECRETED FORM AND DISPLAYS METALLOPROTEASE ACTIVITY.
 CC -1- TISSUE SPECIFICITY: ISOFORM 12L IS EXPRESSED IN PLACENTA AND
 SKELETAL, CARDIAC, AND SMOOTH MUSCLE. ISOFORM 12S SEEMS TO BE
 EXPRESSED ONLY IN PLACENTA OR IN EMBRYO AND FETUS. BOTH FORMS
 WERE EXPRESSED IN SOME TUMOR CELLS LINES. NOT DETECTED IN BRAIN,
 LUNG, LIVER, KIDNEY OR PANCREAS.
 CC -1- DOMAIN: CYS-179 WITHIN THE PROPEPTIDE ACTS AS A CYSTEINE SWITCH,
 BEING ABLE TO CHELATE THE ACTIVE SITE ZINC ION, THUS BLOCKING THE
 ACTIVITY OF THE ENZYME.
 CC -1- DOMAIN: THE CYSTEINE-RICH DOMAIN COULD BE INVOLVED IN TUMOR CELL
 ADHESION.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
 METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 3 SH3-BINDING DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS A DISINTEGRIN DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF023476; AAC08702.1;
 DR EMBL: AF023477; AAC08703.1;
 DR MIM: M12.212;
 DR HSP: P17494; IKST.
 DR PFAM: PF00200; disintegrin; 1.
 DR PFAM: PF01421; Reprolysin; 1.
 DR PFAM: PF01562; Pep_M12B-Propep; 1.
 DR PROSITE: PS50215; ADAM_MEPRO; 1.
 DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
 DR PROSITE: PS50214; DISINTEGRIN_2; 1.
 DR PROSITE: PS00022; EGF_1; FALSE_NEG.
 DR PROSITE: PS0186; EGF_2; FALSE_NEG.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KM Hydrolyase: Metalloproteinase; zinc; Signal; Glycoprotein; Zymogen;
 Transmembrane; EGF-like domain; SH3-binding; Alternative splicing.
 FT SIGNAL. 1 28
 FT PROPEP. 29 207
 FT CHAIN. 208 909
 FT DOMAIN. 208 708
 FT TRANSMEM. 709 729
 FT DOMAIN. 730 909
 FT DOMAIN. 208 416
 FT DOMAIN. 417 512
 FT DOMAIN. 514 649
 FT DOMAIN. 656 688
 FT SITE. 179 179
 FT SITE. 486 488
 FT ACT_SITE. 351 351
 FT METAL. 350 350
 ZINC (CATALYTIC) (POTENTIAL).

FT METAL 354 354 ZINC (CATALYTIC) (POTENTIAL).
 FT METAL 360 360 ZINC (CATALYTIC) (POTENTIAL).
 FT DISULFID 482 495 POTENTIAL.
 FT DISULFID 660 670 BY SIMILARITY.
 FT DISULFID 664 676 BY SIMILARITY.
 FT DISULFID 678 687 BY SIMILARITY.
 FT DOMAIN 834 840 SH3-BINDING (POTENTIAL).
 FT DOMAIN 885 891 SH3-BINDING (POTENTIAL).
 FT DOMAIN 890 896 SH3-BINDING (POTENTIAL).
 FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 452 452 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 705 738 DNOGLTGLVTICGLAAGVYLRKRTILRL -> EAR
 OEAESNRERGGGPEPGESEHASTASTLTL (IN
 ISOFORM 12S).
 FT VARSPLIC 739 909 MISSING (IN ISOFORM 12S).
 SQ SEQUENCE 909 AA; 99642 MW; B74B35DFCBOA70BA CRC64;
 Query Match 12.6%; Score 70.5; DB 1; Length 909;
 Best Local Similarity 27.3%; Pred. No. 13;
 Matches 27; Conservative 10; Mismatches 37; Indels 25; Gaps 6;
 QY 19 ARGICV-PCQINRVGS---GTNAGDLATLATQCSQCPGTALDDG-----VTD 63
 DB 465 AHGCGCEDCQKPRGTACRSDSSNCGLPEFCTGASPHCPANAVYLHDGHCDDVDGYCNG 524
 QY 64 VEDRSAAQCVCCKNFYINGSSPGGEAPGL--OVFAAG 99
 DB 525 ICQTHEDQCV---TLMGPGAKP---APGICFERVNSAG 556
 RESULT 10
 LMGI_MOUSE
 ID LMGI_MOUSE STANDARD; PRT; 1607 AA.
 AC P02468;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE LAMININ GAMMA-1 CHAIN PRECURSOR (LAMININ B2 CHAIN).
 GN LAMC1 OR LAMC-1 OR LAMB-2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88059118; Pubmed=3680290;
 RA Sasaki M., Yamada Y.;
 RT "The laminin B2 chain has a multidomain structure homologous to the
 B1 chain."
 RT J. Biol. Chem. 262:17111-17117(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99000737; Pubmed=3167041;
 RA Durkin M.E., Bartos B.B., Liu S.-H., Phillips S.L., Chung A.E.;
 RT "Primary structure of the mouse laminin B2 chain and comparison with
 the laminin B1."
 RT Biochemistry 27:5198-5204(1988).
 RN [3]
 RP SEQUENCE OF 1-239 FROM N.A.
 RX MEDLINE=88228071; Pubmed=2836421;
 RA Ogawa K., Burbelo P.D., Sasaki M., Yamada Y.;
 RT "The laminin B2 chain promoter contains unique repeat sequences and
 is active in transient transfection."
 RT J. Biol. Chem. 263:8384-8389(1988).
 RN [4]
 RP SEQUENCE OF 1391-1607 FROM N.A.
 RX MEDLINE=85051302; Pubmed=6209134;
 RA Barlow D.P., Green N.M., Kurkinen M., Hogan B.L.M.;
 RT "Sequencing of laminin B chain cDNAs reveals C-terminal regions of
 coiled-coil alpha-helix."

RL EMBL J. 3:2355-2362(1984).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 771-932.
 RX MEDLINE-96196434; PubMed-8648630;
 RA Stetefeld J., Mayer U., Timpl R., Huber R.;
 RT "Crystal structure of three consecutive laminin-type epidermal growth
 factor-like (LE) modules of laminin gamma1 chain harboring the
 RT nidogen binding site."
 RL J. Mol. Biol. 257:644-657(1996).
 RN [6]
 RP STRUCTURE BY NMR OF 824-881.
 RX MEDLINE-96196435; PubMed-8648631;
 RA Baumgartner R., Czisch M., Mayer U., Poeschl E., Huber R.,
 RT Timpl R., Holak T.A.;
 RT "Structure of the nidogen binding LE module of the laminin gamma1
 RT chain in solution."
 RL J. Mol. Biol. 257:658-668(1996).
 CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR. LAMININ
 CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
 CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
 CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
 CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
 CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
 CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE.
 CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
 CC THE GAMMA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ),
 CC LAMININ-2 (MEROSIN), LAMININ-3 (S-LAMININ), LAMININ-4 (S-MEROSIN),
 CC LAMININ-6 (K-LAMININ), AND LAMININ-7 (KS-LAMININ).
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
 CC COMPONENT).
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -1- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
 CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
 CC -1- SIMILARITY: CONTAINS 1 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X05211; CAA28838.1; -
 DR EMBL: J03484; AAA38405.1; -
 DR EMBL: J02930; AAA39408.1; -
 DR EMBL: J03749; AAA39409.1; -
 DR PIR: A28469; MMSB2.
 DR PDB: 1KIO; 20-AUG-97.
 DR PDB: 1TLE; 12-FEB-97.
 DR MGD; MGI:99914; LAMC1.
 DR INTERPRO: IPR000034; -
 DR INTERPRO: IPR000561; -
 DR INTERPRO: IPR001886; -
 DR INTERPRO: IPR002049; -
 DR PFAM: PF000053; laminin_B.1.
 DR PFAM: PF000053; laminin_EGF.10.
 DR PFAM: PF000053; laminin_Nterm.1.
 DR PRINTS: PR00011; EGF_LAMININ.
 DR PROSITE: PS00022; EGF_1; 8.
 DR PROSITE: PS01186; EGF_2; 2.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF.10.
 KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.
 FT CHAIN 1 33
 FT CHAIN 34 1607 LAMININ GAMMA-1 CHAIN.
 FT DOMAIN 34 283 LAMININ N-TERMINAL (DOMAIN VI).
 FT DOMAIN 284 502 4.5 X LAMININ EGF-LIKE REPEATS (DOMAIN
 FT V)
 FT DOMAIN 284 339 LAMININ EGF-LIKE 1.

FT DOMAIN 340 395 LAMININ EGF-LIKE 2.
 FT DOMAIN 396 442 LAMININ EGF-LIKE 3.
 FT DOMAIN 443 492 LAMININ EGF-LIKE 4.
 FT DOMAIN 493 502 LAMININ EGF-LIKE 5 (N-TERMINAL).
 FT DOMAIN 503 687 LAMININ EGF-LIKE 6.
 FT DOMAIN 688 1028 LAMININ EGF-LIKE 7.
 FT DOMAIN 688 721 LAMININ EGF-LIKE 8 (C-TERMINAL).
 FT DOMAIN 722 770 LAMININ EGF-LIKE 9.
 FT DOMAIN 771 825 LAMININ EGF-LIKE 10.
 FT DOMAIN 826 881 LAMININ EGF-LIKE 11.
 FT DOMAIN 882 932 LAMININ EGF-LIKE 12.
 FT DOMAIN 933 980 LAMININ EGF-LIKE 13.
 FT DOMAIN 981 1028 LAMININ EGF-LIKE 14.
 FT DOMAIN 1029 1607 LAMININ EGF-LIKE 15.
 FT DOMAIN 1034 1594 COILED COIL (POTENTIAL).
 FT DISULFID 340 349 BY SIMILARITY.
 FT DISULFID 342 365 BY SIMILARITY.
 FT DISULFID 368 377 BY SIMILARITY.
 FT DISULFID 380 393 BY SIMILARITY.
 FT DISULFID 396 408 BY SIMILARITY.
 FT DISULFID 414 425 BY SIMILARITY.
 FT DISULFID 425 440 BY SIMILARITY.
 FT DISULFID 443 454 BY SIMILARITY.
 FT DISULFID 454 461 BY SIMILARITY.
 FT DISULFID 463 472 BY SIMILARITY.
 FT DISULFID 475 490 BY SIMILARITY.
 FT DISULFID 722 731 BY SIMILARITY.
 FT DISULFID 724 738 BY SIMILARITY.
 FT DISULFID 740 749 BY SIMILARITY.
 FT DISULFID 752 768 BY SIMILARITY.
 FT DISULFID 771 779 BY SIMILARITY.
 FT DISULFID 773 790 BY SIMILARITY.
 FT DISULFID 793 802 BY SIMILARITY.
 FT DISULFID 805 823 BY SIMILARITY.
 FT DISULFID 826 840 BY SIMILARITY.
 FT DISULFID 828 847 BY SIMILARITY.
 FT DISULFID 850 859 BY SIMILARITY.
 FT DISULFID 862 879 BY SIMILARITY.
 FT DISULFID 882 896 BY SIMILARITY.
 FT DISULFID 884 903 BY SIMILARITY.
 FT DISULFID 905 914 BY SIMILARITY.
 FT DISULFID 917 930 BY SIMILARITY.
 FT DISULFID 933 945 BY SIMILARITY.
 FT DISULFID 935 952 BY SIMILARITY.
 FT DISULFID 954 963 BY SIMILARITY.
 FT DISULFID 966 978 BY SIMILARITY.
 FT DISULFID 981 993 BY SIMILARITY.
 FT DISULFID 983 999 BY SIMILARITY.
 FT DISULFID 1001 1010 BY SIMILARITY.
 FT DISULFID 1013 1026 BY SIMILARITY.
 FT DISULFID 1029 1029 INTERCHAIN (PROBABLE).
 FT DISULFID 1032 1032 INTERCHAIN (PROBABLE).
 FT DISULFID 1598 1598 INTERCHAIN (WITH CHAIN
 FT CARBOHYD 58 58 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 132 132 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 574 574 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 648 648 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 1020 1020 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 1105 1105 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 1159 1159 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 1173 1173 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 1203 1203 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 1239 1239 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 1378 1378 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 1393 1393 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 1437 1437 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 216 216 G -> A (IN REF. 3).
 FT CARBOHYD 260 260 E -> D (IN REF. 2).
 FT CARBOHYD 337 337 S -> C (IN REF. 2).
 FT CARBOHYD 447 447 LR -> PS (IN REF. 2).

FT CONFLICT 544 544 D -> Y (IN REF. 2).
 FT CONFLICT 662 662 T -> S (IN REF. 2).
 FT CONFLICT 886 886 MISSING (IN REF. 2).
 FT CONFLICT 1158 1158 MISSING (IN REF. 2).
 FT CONFLICT 1434 1434 V -> A (IN REF. 2).
 FT CONFLICT 1475 1475 R -> K (IN REF. 4).
 FT CONFLICT 1576 1576 D -> N (IN REF. 4).
 SQ SEQUENCE 1607 AA; 177297 MW; 81B7B08E4869F242 CRC64;

Query Match 12.6%; Score 70.5; DB 1; Length 1607;
 Best Local Similarity 22.9%; Pred. No. 22;
 Matches 25; Conservative 9; Mismatches 26; Indels 49; Gaps 5;

OY 23 CVPQINRVSSTNAGDIATLQACST---QCPTGALD-----59
 DB 393 CSPCHSPVGS-----LSTQCDSYGRCSCKRPVGMGDCRCQPGFHSLEAGCRPC 443
 OY 60 -----GTVDFDRSAOCV-----KCKNFY-YNGSPQGEAP 91
 DB 444 SCDLRGSTDECNVETGRGVCKNDVEGFNCERCKPGFFNLESSNPKGCTP 492

RESULT 11

ASP_ANCCA STANDARD; PRT; 424 AA.
 AC Q16937;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE ANCYLOSTOMA SECRETED PROTEIN PRECURSOR.
 GN ASP.
 OS Ancylostoma caninum (Dog hookworm).
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
 ANCYLOSTOMATOIDEA; Ancylostomatidae; Ancylostomatinae; Ancylostoma.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=96215086; PubMed=8636085;
 RA Hawdon J.M., Jones B.F., Hoffman D.R., Hotez P.J.;
 RT "Cloning and characterization of Ancylostoma-secreted protein. A
 RT novel protein associated with the transition to parasitism by
 RT infective hookworm larvae."
 RL J. Biol. Chem. 271:6672-6678(1996).
 CC -1- FUNCTION: ASSOCIATED WITH THE TRANSITION TO PARASITISM BY
 CC INFECTIVE HOOKWORM LARVAE.
 CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
 CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U26187; AAC47001.1; -
 DR HSSP; P04284; ICPE.
 DR INTERPRO: IPR001283; -
 DR PRAM: PR00188; SCP: 1.
 DR PRINTS: PR00837; V5TPYLKE.
 DR PROSITE: PS01009; SCP_AG5_PRI_SC7_1; FALSE_NEG.
 DR PROSITE: PS01010; SCP_AG5_PRI_SC7_2; FALSE_NEG.
 KM signal.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 424 ANCYLOSTOMA SECRETED PROTEIN.
 SQ SEQUENCE 424 AA; 45735 MW; 4B082935D3035F9 CRC64;

Query Match 12.5%; Score 70; DB 1; Length 424;
 Best Local Similarity 26.0%; Pred. No. 7;
 Matches 26; Conservative 4; Mismatches 28; Indels 42; Gaps 4;

OY 8 NGNOPFAANN-----AARGIC-----VPCQINRVSSTN-----AGDIATLQAC 47
 DB 142 NGGGLFAFNSNMVSETTKLCACAYKCGTKLAVSCIVNGYITNQPMWETGACKTGADC 201
 OY 48 ST-----QCPTGALDGGYDVF 65
 DB 202 STYNSGCEDELCTKPGDPVPEFTNOCCPSNTGMTDSVRTF 241

RESULT 12

LMAL_MOUSE STANDARD; PRT; 3084 AA.
 AC P19137;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE LAMININ ALPHA-1 CHAIN PRECURSOR (LAMININ A CHAIN).
 GN LMAL OR LAMA-1 OR LAMA.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RX MEDLINE=89034134; PubMed=3182802;
 RA Sasaki M., Kleinman H.K., Huber H., Deutzmann R., Yamada Y.;
 RT "Laminin, a multidomain protein. The A chain has a unique globular
 RT domain and homology with the basement membrane proteoglycan and the
 RT laminin B chains."
 RL J. Biol. Chem. 263:16536-16544(1988).
 RN [2]
 RP SEQUENCE OF 1-339 FROM N.A.
 RX MEDLINE=88225080; PubMed=3267223;
 RA Hartl L., Oberbauer I., Deutzmann R.;
 RT "The N terminus of laminin A chain is homologous to the B chains."
 RL Eur. J. Biochem. 173:629-635(1988).
 RN [3]
 RP SEQUENCE OF 2538-3084 FROM N.A. AND PARTIAL SEQUENCE.
 RX MEDLINE=89030693; PubMed=3181157;
 RA Deutzmann R., Huber J., Schmeltz K.A., Oberbauer I., Hartl L.;
 RT "Structural study of long arm fragments of laminin. Evidence for
 RT repetitive C-terminal sequences in the A-chain, not present in the B-
 RT chains."
 RL Eur. J. Biochem. 177:35-45(1988).
 CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
 CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
 CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
 CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
 CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
 CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
 CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
 CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
 CC THE ALPHA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ) AND
 CC LAMININ-3 (S-LAMININ).
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
 CC COMPONENT).
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -1- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
 CC -1- SIMILARITY: CONTAINS 17 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
 CC -1- SIMILARITY: CONTAINS 1 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; J04064; AAA39410.1; -

QY 82 NGSGPGGEAP 91
 DB 1147 RGDNPQCSP 1156

RESULT 13
 LMA_DROME STANDARD; PRT; 3712 AA.

AC 000174;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE LAMININ ALPHA CHAIN PRECURSOR.
 GN LANA OR LAMA.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.

RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93049203; PubMed=1425586;
 RA Kuschel-Gullberg M., Garrison K., Mackrell A.J., Fessler L.I.,
 RT "Laminin A chain: expression during Drosophila development and
 genomic sequence."
 RL EMBO J. 11:4519-4527(1992).
 RN (2)
 RP SEQUENCE FROM N.A.
 RX TISSUE-EMBRYO;
 RA MEDLINE=94038678; PubMed=8223265;
 RA Henchcliffe C., Garcia-Alonso L., Tang J., Goodman C.S.;
 RT "Genetic analysis of laminin A reveals diverse functions during
 morphogenesis in Drosophila."
 RL development 118:325-337(1993).

RN (3)
 RP SEQUENCE OF 1762-3712 FROM N.A.
 RX MEDLINE=92078147; PubMed=1744083;
 RA Garrison K., Mackrell A.J., Fessler J.H.;
 RT "Drosophila laminin A chain sequence, interspecies comparison, and
 domain structure of a major carboxyl portion."
 RL J. Biol. Chem. 266:22899-22904(1991).

CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
 IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
 CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
 WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
 CC -1- FUNCTION: DIVERSE FUNCTIONS DURING MORPHOGENESIS IN DROSOPHILA.
 CC COMPLETE LOSS-OF-FUNCTION MUTATIONS LEAD TO LATE EMBRYONIC
 CC LETALITY. CERTAIN PARTIAL LOSS-OF-FUNCTION MUTATIONS GIVE RAISE
 CC TO ESCAPER ADULTS, WHICH HAVE ROUGH EYES ASSOCIATED WITH CHANGES
 CC IN CELL FATE AND PATTERN. MISSHAPEN LEGS AND DEFECTS IN WING
 CC STRUCTURE.
 CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
 CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
 CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
 CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
 CC SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
 CC MEMBRANES (MAJOR COMPONENT).
 CC -1- TISSUE SPECIFICITY: NEWLY FORMED MESODERM AND LATER PROVENTRICLY
 CC EXPRESSED IN HEMOCYTES, WHICH ALSO SYNTHESIZE COLLAGEN IV.
 CC -1- DEVELOPMENTAL STAGE: DURING MORPHOGENESIS, MOSTLY IN EMBRYO
 CC DEVELOPMENT AT 10-12 HOURS.
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -1- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
 CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
 CC -1- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILAR TO LAMININ DOMAIN IV (DOMAIN IV) IS NOT
 CC SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: M96388; AAA28662.1; -
 CC EMBL: L07288; AAC37178.1; -
 CC EMBL: M75882; AAA28661.1; -
 CC HSSP: P02468; 1TLE.
 CC DR FLYBASE; FBgn0002526; LANA.
 CC DR INTERPRO; IPR000034; -
 CC DR INTERPRO; IPR000561; -
 CC DR INTERPRO; IPR001791; -
 CC DR INTERPRO; IPR001886; -
 CC DR INTERPRO; IPR002049; -
 CC DR PFAM; PF00052; laminin_B; 1.
 CC DR PFAM; PF00053; laminin_EGF; 20.
 CC DR PFAM; PF00054; laminin_G; 5.
 CC DR PFAM; PF00055; laminin_Nterm; 1.
 CC DR PRINTS; PR00011; EGF-LAMININ.
 CC DR PROSITE; PS00022; EGF_1; 17.
 CC DR PROSITE; PS01166; EGF_2; 5.
 CC DR PROSITE; PS01246; LAMININ_TYPE_EGF; 19.
 CC KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 CC Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 CC FT SIGNAL 1 22
 CC FT CHAIN 23 3712
 CC FT DOMAIN 25 272
 CC FT DOMAIN 273 815

FT 273 332 LAMININ EGF-LIKE 1.
 FT 333 402 LAMININ EGF-LIKE 2.
 FT 403 447 LAMININ EGF-LIKE 3.
 FT 448 494 LAMININ EGF-LIKE 4.
 FT 495 540 LAMININ EGF-LIKE 5.
 FT 541 586 LAMININ EGF-LIKE 6.
 FT 587 631 LAMININ EGF-LIKE 7.
 FT 632 676 LAMININ EGF-LIKE 8.
 FT 677 731 LAMININ EGF-LIKE 9.
 FT 732 784 LAMININ EGF-LIKE 10.
 FT 785 815 LAMININ EGF-LIKE 11 (INCOMPLETE).
 FT 816 1374 LAMININ IV.
 FT 817 1374 LAMININ EGF-LIKE REPEATS (DOMAIN
 FT 1375 1420 I II B).
 FT 1421 1465 LAMININ EGF-LIKE 12.
 FT 1466 1513 LAMININ EGF-LIKE 13.
 FT 1514 1564 LAMININ EGF-LIKE 14.
 FT 1565 1574 LAMININ EGF-LIKE 15.
 FT 1575 1775 LAMININ EGF-LIKE 16 (N-TERMINAL).
 FT 1776 2111 LAMININ DOMAIN IV (DOMAIN IV).
 FT 2112 2111 6.5 X LAMININ EGF-LIKE REPEATS (DOMAIN
 FT 2112 2111 III A).
 FT 1776 1808 LAMININ EGF-LIKE 16 (C-TERMINAL).
 FT 1809 1858 LAMININ EGF-LIKE 17.
 FT 1859 1916 LAMININ EGF-LIKE 18.
 FT 1917 1969 LAMININ EGF-LIKE 19.
 FT 1970 2016 LAMININ EGF-LIKE 20.
 FT 2017 2063 LAMININ EGF-LIKE 21.
 FT 2064 2111 LAMININ EGF-LIKE 22.
 FT 2112 2111 LAMININ II AND I.
 FT 2112 2697 5 X LAMININ G-LIKE REPEATS (DOMAIN G).
 FT 2698 2862 LAMININ G-LIKE 1.
 FT 2863 3048 LAMININ G-LIKE 2.
 FT 3049 3223 LAMININ G-LIKE 3.
 FT 3270 3296 POLY-THR.
 FT 3297 3528 LAMININ G-LIKE 4.
 FT 3529 3712 LAMININ G-LIKE 5.
 FT 3713 3712 LAMININ G-LIKE 6.
 FT 2178 2249 COILED COIL (POTENTIAL).
 FT 2301 2321 COILED COIL (POTENTIAL).
 FT 2376 2450 COILED COIL (POTENTIAL).
 FT 2541 2676 COILED COIL (POTENTIAL).
 FT 273 282 DISULFID


```

RESULT 15
TSP2_MOUSE STANDARD; PRT; 1172 AA.
ID TSP2_MOUSE
AC 003350:
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE THROMBOSPONDIN 2 PRECURSOR.
GN THBS2 OR TSP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92147683; Pubmed=1371115;
RA Laherty C.D., O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,
RA Dixit V.M.;
RT "Characterization of mouse thrombospondin 2 sequence and expression
RT during cell growth and development."
RL J. Biol. Chem. 267:3274-3281(1992).
RN [2]
RP SEQUENCE OF 1-873 FROM N.A.
RX MEDLINE=91302287; Pubmed=1712771;
RA Bornstein P., O'Rourke K., Wikstrom K., Wolf F.W., Katz R., Li P.,
RA Dixit V.M.;
RT "A second, expressed thrombospondin gene (Thbs2) exists in the mouse
RT genome."
RL J. Biol. Chem. 266:12821-12824(1991).
RN [3]
RP FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
RP CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
RP LAMININ AND TYPE V COLLAGEN.
CC -1 SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS.
CC -1 SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1 SIMILARITY: CONTAINS 1 WFEC DOMAIN.
CC -1 SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -1 SIMILARITY: CONTAINS 3 TYPE-1 TSP REPEATS AND 7 TYPE-3 TSP REPEATS
CC (WHICH BIND CALCIUM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L07803; AAA53064.1; -
DR EMBL: M64866; AAA40432.1; -
DR PIR: A42587; A42587.
DR PIR: A39851; A39851.
DR HSSP: P00740; 11XA.
DR MGD: MGI:98738; THBS2.
DR INTERPRO: IPR000561; -
DR INTERPRO: IPR000884; -
DR INTERPRO: IPR001007; -
DR PFAM: PF00008; EGF_2;
DR PFAM: PF00090; tsp_1; 3.
DR PROSITE: PS00022; EGF_1; FALSE_NEG.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS00092; TSP1; 3.
DR PROSITE: PS01208; WFEC; 1.
DR Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT SIGNAL 1 18
FT CHAIN 19 1172
FT DOMAIN 19 232
FT DOMAIN 318 375
FT DOMAIN 381 548
FT DOMAIN 549 692
FT DOMAIN 725 952
FT DOMAIN 953 1172

```

```

FT REPEAT 381 432 TSP TYPE-1 1.
FT REPEAT 437 493 TSP TYPE-1 2.
FT REPEAT 494 548 TSP TYPE-1 3.
FT DOMAIN 549 589 EGF-LIKE 1.
FT DOMAIN 590 647 EGF-LIKE 2.
FT DOMAIN 648 692 EGF-LIKE 3.
FT REPEAT 725 760 TSP TYPE-3 1.
FT REPEAT 761 783 TSP TYPE-3 2.
FT REPEAT 784 819 TSP TYPE-3 3.
FT REPEAT 820 842 TSP TYPE-3 4.
FT REPEAT 843 880 TSP TYPE-3 5.
FT REPEAT 881 916 TSP TYPE-3 6.
FT REPEAT 917 952 TSP TYPE-3 7.
FT SITE 928 930 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 266 266 INTERCHAIN (PROBABLE).
FT DISULFID 270 270 INTERCHAIN (PROBABLE).
FT DISULFID 553 564 BY SIMILARITY.
FT DISULFID 558 574 BY SIMILARITY.
FT DISULFID 577 588 BY SIMILARITY.
FT DISULFID 594 610 BY SIMILARITY.
FT DISULFID 601 619 BY SIMILARITY.
FT DISULFID 622 646 BY SIMILARITY.
FT DISULFID 652 665 BY SIMILARITY.
FT DISULFID 659 678 BY SIMILARITY.
FT DISULFID 680 691 BY SIMILARITY.
FT CARBOHYD 151 151 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 584 584 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1069 1069 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1172 AA; 129911 MW; 7C8E4E859822AB CRC64;

```

```

Query Match 12.38; Score 69; DB 1; Length 1172;
Best Local Similarity 26.7%; Pred. No. 23;
Matches 20; Conservative 9; Mismatches 22; Indels 24; Gaps 4;
QY 23 CVPQINRVSGSTNAGDPLATLQCSITQCPRTGLDGVTV--FDRSAGCVKCKPRFY 80
DB 574 CGSCPVGFLGNGTCELD---DCAV-----VYDICSFTNKAQRCVMTNNGFH 618
QY 81 -----YNGSGPOG 88
DB 619 CLPCPPRRKGNQPRG 633

```

Search completed: March 6, 2001, 12:54:46
 Job time: 403 sec

THIS PAGE BLANK (USPTO)

APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Immediate Early Protein From Kaposi's
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
TITLE OF INVENTION: Encoding Same And Uses Thereof
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728.323A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/52268/JPM/MS/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3489
US-08-728-323A-1

Query Match 10.8% Score 34; DB 2; Length 3489;
Best Local Similarity 44.4% Pred. No. 0.15;
Matches 136; Conservative 0; Mismatches 170; Indels 0; Gaps 0;
OY 5 ccgctcaggaggaagctaagttaacagccttcgcagcaataaagtctgtagagta 64
DB 2106 CTGCTCTGCTCATCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTCATC 2047
OY 65 tatgtaccatgcgaataaacaagagtaggctctgtaccatgcagagtaactagta 124
DB 2046 CTGCTCTGCTCATCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTCATC 1987
OY 125 cttagcacaacaatgaagtaactcaagtgctactgacacgacactgaatgaatga 184
DB 1986 CTGCTGCTGCTCATCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCATC 1927
OY 185 cagatgtttttagatagatcagccgacagatgtgttaaatgaataaacttaactata 244
DB 1926 CTGCTGCTGCTCATCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCATC 1867
OY 245 atggtgttctccctcaggatgaagctcctgagccttgaagtttgcagcgtgagtgccg 304
DB 1866 CTGCTGCTGCTCATCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCATC 1807
OY 305 ctgcag 310
DB 1806 CTGCTG 1801

RESULT 3
US-08-770-379-20
Sequence 20, Application US/08770379
Patent No. 5849564
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770.379
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52342
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 32207 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-770-379-20

Query Match 10.8% Score 34; DB 2; Length 32207;
Best Local Similarity 44.4% Pred. No. 0.41;
Matches 136; Conservative 0; Mismatches 170; Indels 0; Gaps 0;
OY 5 ccgctcaggaggaagctaagttaacagccttcgcagcaataaagtctgtagagta 64
DB 19891 CTGCTCTGCTCATCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTCATC 19950
OY 65 tatgtaccatgcgaataaacaagagtaggctctgtaccatgcagagtaactagta 124
DB 19951 CTGCTGCTGCTCATCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCATC 20010
OY 125 cttagcacaacaatgaagtaactcaagtgctactgacacgacactgaatgaatga 184
DB 20011 CTGCTGCTGCTCATCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCATC 20070
OY 185 cagatgtttttagatagatcagccgacagatgtgttaaatgaataaacttaactata 244
DB 20071 CTGCTGCTGCTCATCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCATC 20130
OY 245 atggtgttctccctcaggatgaagctcctgagccttgaagtttgcagcgtgagtgccg 304
DB 20131 CTGCTGCTGCTCATCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCATC 20190
OY 305 ctgcag 310
DB 20191 CTGCTG 20196

```

RESULT 4
US-08-469-802B-3/C
; Sequence 3, Application US/08469802B
; Patent No. 5741645
; GENERAL INFORMATION:
; APPLICANT: OTT, Harry T.
; APPLICANT: Rannum, Laura P.W.
; APPLICANT: Chung, Ming-Yi
; APPLICANT: Zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5741645
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Muelting, Raasch, Gebhardt & Schwappach, P.A.
; STREET: 119 No. 5741645th Fourth Street, Suite 203
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,802B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Muelting, Ann M.
; REGISTRATION NUMBER: 33,977
; REFERENCE/DOCKET NUMBER: 110.00030101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1225
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-08-469-802B-3

Query Match 10.6%; Score 33.6; DB 1; Length 234;
Best Local Similarity 47.2%; Pred. No. 0.055;
Matches 102; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 95 gctctgagcaacatgcaagtgactagctactttagccacacaaatgcagtgactcagtgctc 154
DB 216 gctgaggtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 157
QY 155 ctactgagcactgacactgatgatgagatgacagatgctttagatagatcagccagcagct 214
DB 156 ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 97
QY 215 gtgttaaatgcaaacctaaccttactataatggtggttctctcctcaggggtgaagctctg 274
DB 96 ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 37
QY 275 gctctcaggttctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 310
DB 36 ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1

```

```

; GENERAL INFORMATION:
; APPLICANT: OTT, Harry T.
; APPLICANT: Rannum, Laura P.W.
; APPLICANT: Chung, Ming-Yi
; APPLICANT: Zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5834183
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Muelting, Raasch, Gebhardt & Schwappach, P.A.
; STREET: P.O. Box 581415
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55458-1415
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/267,803B
; FILING DATE: 28-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McCormack, Myra H.
; REGISTRATION NUMBER: 36,602
; REFERENCE/DOCKET NUMBER: 110.00030120
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1228
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-08-267-803B-3

Query Match 10.6%; Score 33.6; DB 2; Length 234;
Best Local Similarity 47.2%; Pred. No. 0.055;
Matches 102; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 95 gctctgagcaacatgcaagtgactagctactttagccacacaaatgcagtgactcagtgctc 154
DB 216 gctgaggtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 157
QY 155 ctactgagcactgacactgatgatgagatgacagatgctttagatagatcagccagcagct 214
DB 156 ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 97
QY 215 gtgttaaatgcaaacctaaccttactataatggtggttctctcctcaggggtgaagctctg 274
DB 96 ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 37
QY 275 gctctcaggttctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 310
DB 36 ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1

RESULT 6
US-09-253-691-3/C
; Sequence 3, Application US/09253691
; Patent No. 6124100
; GENERAL INFORMATION:
; APPLICANT: Dong KYU JIN
; TITLE OF INVENTION: Diagnostic Method and Kit for Neuropsychiatric Diseases
; TITLE OF INVENTION: Using Trinucleotide Repeats Sequence
; FILE REFERENCE: 1942/36
; CURRENT APPLICATION NUMBER: US/09/253,691

```

```

: CURRENT FILING DATE: 1999-02-22
: EARLIER APPLICATION NUMBER: KR 98-6,278
: EARLIER FILING DATE: 1996-02-26
: NUMBER OF SEQ ID NOS: 3
: SOFTWARE: wordperfect 6.1/Windows
: SEQ ID NO 3
: LENGTH: 397
: TYPE: DNA
: ORGANISM: human
: US-09-253-691-3

```

```

Query Match      10.6%; Score 33.4; DB 3; Length 397;
Best Local Similarity 47.0%; Pred. No. 0.082;
Matches 103; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

```

```

Oy 92 tagctctggtaccatgcaggtgacttagctacttagccacacagcaggtactcagt 151
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 364 TAGGTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 305
Oy 152 gtccactgacacgtgactgatatgagatgtttttatagatcagccgac 211
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 304 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 245
Oy 212 agtgtltaaatgcaaaccttaacttacttaatagtgtgttctcccaagggtgaagctc 271
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 244 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 185
Oy 272 ctggccctcaggttttctgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 310
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 146

```

```

RESULT 7
US-08-149-097D-23
: Sequence 23, Application US/08149097D
: Patent NO. 5874236
: GENERAL INFORMATION:
: APPLICANT: Harpold, Michael
: APPLICANT: Ellis, Steven
: APPLICANT: Williams, Mark
: APPLICANT: Feldman, Daniel
: APPLICANT: McCue, Ann
: APPLICANT: Brenner, Robert
: TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
: NUMBER OF SEQUENCES: 40
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Brown, Martin, Haller & McClain
: STREET: 1660 Union Street
: CITY: San Diego
: STATE: California
: COUNTRY: USA
: ZIP: 92101-2926
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/149,097D
: FILING DATE: 05-NOV-1993
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/105,536
: FILING DATE: 11-AUG-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/US92/06903
: FILING DATE: 14-AUG-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/914,231
: FILING DATE: 13-JUL-1992
: PRIOR APPLICATION DATA:

```

```

: APPLICATION NUMBER: US 07/868,354
: FILING DATE: 10-APR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/745,206
: FILING DATE: 15-AUG-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/620,250
: FILING DATE: 30-NOV-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/482,384
: FILING DATE: 20-FEB-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/603,751
: FILING DATE: 04-APR-1989
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/US89/01408
: FILING DATE: 04-APR-1989
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/176,899
: FILING DATE: 04-APR-1988
: ATTORNEY/AGENT INFORMATION:
: NAME: Seidman, Stephanie L.
: REGISTRATION NUMBER: 33,779
: REFERENCE/DOCKET NUMBER: 6362-55038
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 238-0062
: TELEFAX: (619) 238-0062
: INFORMATION FOR SEQ ID NO: 23:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7791 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 240..7037
: OTHER INFORMATION: /product= "Alpha1a-2 subunit of
: US-08-149-097D-23

```

```

Query Match      10.3%; Score 32.6; DB 2; Length 7791;
Best Local Similarity 50.3%; Pred. No. 0.6;
Matches 80; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
Oy 94 ggcctgtaccatgacgaggtgacttagctacttagccacacaaatgagtaactcagtg 153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 812 GACGCTGAGGACAGTTCAGTGCAGTGCAGGCGCGCTCAGCTGAGTGTGTGGAATCCCAAGTTT 871
Oy 154 cctactgacactgacactgagatgagatgacagatgtttttagatagatcagccgacag 213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 872 ACAAGTCTCTCGAAGTGCATGATGAAAGCGATGATCCCTTTGCTGCGAGATCGGCTTCT 931
Oy 214 tgtltaaatgcaaaccttaacttacttaataatggtgtgt 252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 932 CCTATTTTTTCGAATCCTTATTTTTTCGAATCATATAGGGTT 970

```

```

RESULT 8
US-08-949-386-23
: Sequence 23, Application US/08949386
: Patent NO. 6090623
: GENERAL INFORMATION:
: APPLICANT: Harpold, Michael
: APPLICANT: Ellis, Steven
: APPLICANT: Williams, Mark
: APPLICANT: McCue, Ann
: APPLICANT: Gillespie, Allison
: TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
: NUMBER OF SEQUENCES: 38
: CORRESPONDENCE ADDRESSES:

```


ADDRESSEE: Brown, Martin, Haller & McClain
 STREET: 1660 Union Street
 CITY: San Diego
 STATE: California
 COUNTRY: US
 ZIP: 92101
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/949,386
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/290,012
 FILING DATE: 11-AUG-1994
 APPLICATION NUMBER: 08/149,097
 FILING DATE: 5-NOV-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/105,536
 FILING DATE: 11-AUG-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Seidman, Stephanie L.
 REGISTRATION NUMBER: 33,779
 REFERENCE/DOCKET NUMBER: 519808
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 238-0999
 TELEFAX: (619) 238-0062
 INFORMATION FOR SEQ ID NO: 23:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 7791 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 237..7037
 OTHER INFORMATION: /standard_name="Alpha-1A-2"
 US-08-949-386-23
 Query Match 10.3%; Score 32.6; DB 3; Length 7791;
 Best Local Similarity 50.3%; Pred. No. 0.6;
 Matches 80; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
 TITLE OF INVENTION: METHODS
 NUMBER OF SEQUENCES: 38
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Brown, Martin, Haller & McClain
 STREET: 1660 Union Street
 CITY: San Diego
 STATE: California
 COUNTRY: US
 ZIP: 92101
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/450,562
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/404,950
 FILING DATE: 13-MAR-1995
 APPLICATION NUMBER: 08/336,257
 FILING DATE: 7-NOV-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/314,083
 FILING DATE: 28-SEPT-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/311,363
 FILING DATE: 23-SEPT-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/290,012
 FILING DATE: 11-AUG-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/223,305
 FILING DATE: 4-APR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/193,078
 FILING DATE: 07-FEB-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/149,097
 FILING DATE: 5-NOV-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/105,536
 FILING DATE: 11-AUG-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/914,231
 FILING DATE: 13-JULY-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/868,354
 FILING DATE: 10-APR-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/US92/06903
 FILING DATE: 14-AUG-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/745,206
 FILING DATE: 15-AUG-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/620,250
 FILING DATE: 30-NOV-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/603,751
 FILING DATE: 08-NOV-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/482,384
 FILING DATE: 02-FEB-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/US89/01408
 FILING DATE: 04-APR-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/176,899
 FILING DATE: 04-APR-1988
 ATTORNEY/AGENT INFORMATION:

RESULT 9
 US-08-450-562-23
 Sequence 23, Application US/08450562
 Patent No. 6096514
 GENERAL INFORMATION:
 APPLICANT: Harpold, Michael
 APPLICANT: Ellis, Steven
 APPLICANT: Williams, Mark
 APPLICANT: McCue, Ann
 APPLICANT: Gillespie, Allison
 APPLICANT: Seidman, Daniel
 APPLICANT: Brenner, Robert

Db 94 ggcctgtaccacatgcagtgactgactttagccacacatgcagtgactgactgct 153
 Db 812 GACCGTCAGGCGCAGTTCGAGCTCGCGCCGCTCAAGCTGCTGGAATCCCAAGTTT 871
 Oy 154 cctactgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatg 213
 Db 872 ACAAGTCGCTGGAAGTCATGATGAAGCGATGATCCCTTTCGAGATCGGCTCTCT 931
 Oy 214 tctgttaaatgcaaacctacttactcataatggtgt 252
 Db 932 CCTATTTTTCGATCCTTATTTTTCGATCATAGGTT 970

```

1  PRIOR APPLICATION DATA:
2  APPLICATION NUMBER: US 07/914,231
3  FILING DATE: 13-JUL-1992
4  PRIOR APPLICATION DATA:
5  APPLICATION NUMBER: US 07/868,354
6  FILING DATE: 10-APR-1992
7  PRIOR APPLICATION DATA:
8  APPLICATION NUMBER: US 07/745,206
9  FILING DATE: 15-AUG-1991
10 PRIOR APPLICATION DATA:
11 APPLICATION NUMBER: US 07/620,250
12 FILING DATE: 30-NOV-1990
13 PRIOR APPLICATION DATA:
14 APPLICATION NUMBER: US 07/482,384
15 FILING DATE: 20-FEB-1990
16 PRIOR APPLICATION DATA:
17 APPLICATION NUMBER: US 07/603,751
18 FILING DATE: 04-APR-1989
19 PRIOR APPLICATION DATA:
20 APPLICATION NUMBER: WO PCT/US89/01408
21 FILING DATE: 04-APR-1985
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: US 07/176,899
24 FILING DATE: 04-APR-1988
25 ATTORNEY/AGENT INFORMATION:
26 NAME: Seidman, Stephanie L.
27 REGISTRATION NUMBER: 33,779
28 REFERENCE/DOCKET NUMBER: 6362-55038
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: (619) 238-0999
31 TELEFAX: (619) 238-0062
32 INFORMATION FOR SEQ ID NO: 22:
33 SEQUENCE CHARACTERISTICS:
34 LENGTH: 7808 base pairs,
35 TYPE: nucleic acid
36 STRANDEDNESS: single
37 TOPOLOGY: unknown
38 MOLECULE TYPE: DNA (genomic)
39 FEATURE:
40 NAME/KEY: CDS
41 LOCATION: 240..7769
42 OTHER INFORMATION: /product="Alpha1A-1 subunit of
43 OTHER INFORMATION: human calcium channel"
44 US-08-149-097D-22
45
46 Query Match 10.3%; Score 32.6; DB 2; Length 7808;
47 Best Local Similarity 50.3%; Pred. No. 0.6;
48 Matches 80; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
49
50 Oy 94 ggcctgtgaccgaatgcagcgtgacttagctacttagccacacaaatgcagtaactcaagt 153
51 Db 812 GACCGTGGAGGGCAGCTTCAGAGCTCCGGCCCGCTCAAGCTGGTGTGCGAATCCCAAGTTT 871
52 Oy 154 cctactgcacgtgcactgataatgagtgagtgacagatgtttttgatagatcagccgcaag 213
53 Db 872 ACAAGTCGTCTCGAAGTGCATCATGATAAGGCCATGATCCCTTTTGCTGCAATGCGCTCTCT 931
54 Oy 214 tgtgtaaatgcaaacctaacttacttactataaagtgtgt 252
55 Db 932 CCTATTTTTCGAATCCTATTATTTTTCGAATCATATAGGTT 970
56
57 RESULT 11
58 US-08-949-386-22
59 Sequence 22, Application US/08949386
60 Patent No. 6090523
61 GENERAL INFORMATION:
62 APPLICANT: Harpold, Michael
63 APPLICANT: Ellis, Steven
64 APPLICANT: Williams, Mark
65 APPLICANT: McCue, Ann
66 APPLICANT: Gillespie, Allison

```

```

1  TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
2  TITLE OF INVENTION: METHODS
3  NUMBER OF SEQUENCES: 38
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE: Brown, Martin, Haller & McClain
6  STREET: 1660 Union Street
7  CITY: San Diego
8  STATE: California
9  COUNTRY: US
10 ZIP: 92101
11
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: Floppy disk
14 COMPUTER: IBM PC compatible
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16 SOFTWARE: Patent Release #1.0, Version #1.25
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER: US/08/949,386
19 FILING DATE:
20 CLASSIFICATION:
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: US/08/290,012
23 FILING DATE: 11-AUG-1994
24 APPLICATION NUMBER: 08/149,097
25 FILING DATE: 5-NOV-1993
26 PRIOR APPLICATION DATA:
27 APPLICATION NUMBER: 08/105,536
28 FILING DATE: 11-AUG-1993
29 ATTORNEY/AGENT INFORMATION:
30 NAME: Seidman, Stephanie L.
31 REGISTRATION NUMBER: 33,779
32 REFERENCE/DOCKET NUMBER: 519808
33 TELECOMMUNICATION INFORMATION:
34 TELEPHONE: (619) 238-0999
35 TELEFAX: (619) 238-0062
36 INFORMATION FOR SEQ ID NO: 22:
37 SEQUENCE CHARACTERISTICS:
38 LENGTH: 7808 base pairs
39 TYPE: nucleic acid
40 STRANDEDNESS: double
41 TOPOLOGY: linear
42 MOLECULE TYPE: DNA (genomic)
43 FEATURE:
44 NAME/KEY: CDS
45 LOCATION: 237..7769
46 OTHER INFORMATION: /standard_name="Alpha-1a-1"
47
48 US-08-949-386-22
49
50 Query Match 10.3% Score 32.6; DB 3; Length 7808;
51 Best Local Similarity 50.3%; Fred. No. 0.6;
52 Matches 80; Conservative 0; Mismatches 79; Indels 0; Gaps 0.
53
54 94 ggcctcgtgaccacgaagcagtgactagctacttagccacacacgaagtcagtaactgctg 153
55 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
56 Db 812 GACCGCTGAGGCGAGTTCGAGTCTCGCGGCCCTCAAGCTGGTGTCTGCAATCCGAGTTT 871
57 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
58 154 cctactgacactgcacttgatgatgagtgacagatglttttcatagatcagccgcacag 213
59 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
60 Db 872 ACAAGTCGTCCTGAGAGTGCATCATGAAGCGCATGATCCCTTTGGCGCAGCATGGGCTCCT 931
61 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
62 214 tctgttaaatgcacaaccttaactttactataatagtgtgt 252
63 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
64 Db 932 CCTATTTTTGGCAATCCTTATTTTTGGCAATCATAGGGTT 970
65 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
66
67 RESULT 12
68 US-08-450-562-22
69 Sequence 22, Application US/08450562
70 Patent No. 6096514
71 GENERAL INFORMATION:
72 APPLICANT: Harpold, Michael
73 APPLICANT: Ellis, Steven
74 APPLICANT: Williams, Mark

```

APPLICANT: McGue, Ann
APPLICANT: Gillespie, Alison
APPLICANT: Feldman, Daniel
APPLICANT: Brenner, Robert
TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
METHODS
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: California
COUNTRY: US
ZIP: 92101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/450,562
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/404,950
FILING DATE: 13-MAR-1995
APPLICATION NUMBER: 08/336,257
FILING DATE: 7-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/314,083
FILING DATE: 28-SEPT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/311,363
FILING DATE: 23-SEPT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/290,012
FILING DATE: 11-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/223,305
FILING DATE: 4-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/193,078
FILING DATE: 07-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/149,097
FILING DATE: 5-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/105,536
FILING DATE: 11-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/914,231
FILING DATE: 13-JULY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/868,354
FILING DATE: 10-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/06903
FILING DATE: 14-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/745,206
FILING DATE: 15-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/620,250
FILING DATE: 30-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/603,751
FILING DATE: 08-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/482,384
FILING DATE: 02-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US89/01408
FILING DATE: 04-APR-1989

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2001, 15:12:10 ; Search time 947.87 Seconds
(without alignments)
2336.150 Million cell updates/sec

Title: US-09-196-161d-9

Perfect score: 316
Sequence: 1 ggatccgcgcagggagaagc.....tgctgcgcctgcagaattc 316

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq Length: 0

Maximum DB seq Length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_esthum1:*
43: em_esthum2:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estpl1:*
70: em_estpl2:*
71: em_estpl3:*
72: em_estpl4:*
73: em_estpl5:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_esta1:*
88: gb_esta2:*
89: gb_esta3:*
90: gb_esta4:*
91: gb_esta5:*
92: gb_esta6:*
93: gb_esta7:*
94: gb_esta8:*
95: gb_esta9:*
96: gb_esta10:*
97: gb_esta11:*
98: gb_esta12:*
99: gb_esta13:*
100: gb_esta14:*
101: gb_esta15:*
102: gb_esta16:*
103: gb_esta17:*
104: gb_esta18:*
105: gb_esta19:*
106: gb_esta20:*
107: gb_esta21:*
108: gb_esta22:*
109: gb_esta23:*
110: gb_esta24:*
111: gb_esta25:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estcom1:*
116: em_estcom2:*

117: em_estp16:*
 118: em_estp17:*
 119: em_estp18:*
 120: em_estro14:*
 121: em_estro15:*
 122: em_estro16:*
 123: em_estro17:*
 124: em_estro18:*
 125: em_estro19:*
 126: gb_est58:*
 127: gb_est59:*
 128: gb_est60:*
 129: gb_est61:*
 130: gb_est62:*
 131: gb_est63:*
 132: gb_est64:*
 133: gb_est65:*
 134: gb_est66:*
 135: gb_est67:*
 136: gb_est68:*
 137: gb_est69:*
 138: gb_est70:*
 139: gb_est71:*
 140: gb_est72:*
 141: gb_est73:*
 142: gb_est74:*
 143: gb_est75:*
 144: gb_est76:*
 145: gb_est77:*
 146: gb_est78:*
 147: gb_est79:*
 148: gb_est80:*
 149: gb_est81:*
 150: gb_est82:*
 151: gb_est83:*
 152: gb_est84:*
 153: gb_est85:*
 154: gb_est86:*
 155: gb_est87:*
 156: gb_est88:*
 157: gb_est89:*
 158: gb_est90:*
 159: gb_est91:*
 160: gb_est92:*
 161: gb_est93:*
 162: gb_est94:*
 163: gb_est95:*
 164: gb_est96:*
 165: gb_est97:*
 166: gb_est98:*
 167: gb_est99:*
 168: gb_est100:*
 169: gb_est101:*
 170: gb_est102:*
 171: gb_est103:*
 172: gb_est104:*
 173: gb_est105:*
 174: gb_est106:*
 175: gb_est107:*
 176: gb_est108:*
 177: gb_est109:*
 178: gb_est110:*
 179: gb_est111:*
 180: gb_est112:*
 181: gb_est113:*
 182: gb_est114:*
 183: gb_est115:*
 184: gb_est116:*
 185: gb_est117:*
 186: gb_est118:*
 187: gb_est119:*
 188: gb_est120:*
 189: gb_est121:*

190: gb_gss25:*
 191: gb_gss26:*
 192: gb_gss27:*
 193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	43	13.6	412	193	FR0029465	AL025834 Fugu rubripes
2	37.8	12.0	619	193	FR0006944	290754 F. rubripes
3	37.6	11.9	492	90	AM465310	AM465310 BP23001B8
4	37.4	11.8	805	192	CNS04RW2	AL304427 Tetradon
5	37	11.7	526	178	A2312601	A2312601 IM0028104
6	37	11.7	571	172	A0919110	A0919110 RPCT-23-2
7	37	11.7	610	160	A0524063	A0524063 HS_5232.B
8	36.8	11.6	621	161	A0564861	A0564861 HS_5361.B
9	36.6	11.6	446	105	BE224238	BE224238 kp36d09.y
10	35.6	11.3	454	190	AF046361	AF046361 Mus muscu
11	35.6	11.3	928	190	CNS00R40	AL070039 Drosophila
12	35.6	11.3	978	192	CNS04OR1	AL269254 Tetradon
13	35.4	11.2	803	175	A2101123	A2101123 RPCT-23-4
14	35.2	11.1	576	11	AA754683	AA754683 vut0610.r
15	35.2	11.1	849	190	CNS0115X	AL099855 Drosophila
16	35.2	11.1	906	191	CNS021LL	AL199074 Tetradon
17	35	11.1	382	134	BE027589	BE027589 Elestrea47
18	35	11.1	399	24	AI755783	AI755783 Elestrea20
19	35	11.1	614	151	AQ257640	AQ257640 nbx0018D
20	35	11.1	667	192	CNS03YXK	AL266285 Tetradon
21	35	11.1	1003	192	CNS040UZ	AL303092 Tetradon
22	34.8	11.0	369	190	AG024045	AG024045 Oryza sat
23	34.8	11.0	770	135	BE744403	BE744403 60157651
24	34.8	11.0	946	15	AI069309	AI069309 mgae0006c
25	34.4	10.9	625	173	AC055585	AC055585 LERAE83TR
26	34.2	10.8	831	111	BE705571	BE705571 SC01_03b1
27	34.2	10.8	946	190	AF011110	AF011110 Homo sapi
28	34.2	10.8	1076	193	CNS05RXN	AL338180 Tetradon
29	34	10.8	326	2	AA143080	AA143080 z069b01.r
30	34	10.8	335	140	D66671	D66671 CEK1120E2R
31	34	10.8	387	172	AQ0907407	AQ0907407 GSSRC0843
32	34	10.8	484	5	AA314486	AA314486 EST186534
33	34	10.8	513	151	AQ316331	AQ316331 RPCT111-10
34	34	10.8	537	177	A2264209	A2264209 RPCT-23-1
35	34	10.8	572	192	CNS0435N	AL273200 Tetradon
36	34	10.8	863	192	CNS04HUF	AL291408 Tetradon
37	34	10.8	970	192	CNS03H6V	AL243904 Tetradon
38	33.8	10.7	343	177	A2250232	A2250232 RPCT-23-5
39	33.8	10.7	649	110	BE581913	BE581913 kp36c10.y
40	33.8	10.7	652	110	BE581861	BE581861 kp36c02.y
41	33.6	10.6	445	19	AI356219	AI356219 yv65e03.x
42	33.6	10.6	507	5	AA318257	AA318257 EST20320
43	33.6	10.6	648	18	AI257238	AI257238 LP05446.5
44	33.6	10.6	987	97	AM940443	AM940443 CH12043.3
45	33.6	10.6	1101	190	CNS00R66	AL077346 Drosophila

ALIGNMENTS

RESULT 1
 FR0029465/c 412 bp DNA
 LOCUS Fugu rubripes GSS sequence, clone 047B13aC11, genomic survey
 DEFINITION
 ACCESSION AL025834
 VERSION AL025834.1 GI:3263177
 KEYWORDS GSS; genome survey sequence.
 SOURCE Fugu rubripes.

/strain="Pilariform larvae obtained from humans"
 /db_xref="taxon:6248"
 /clone_lib="TBN95FM-SSFM"
 /lab_host="XL-1 Blue MRF" (Stratagene)
 /note="Vector: Lambda Uni-ZAP XR (Stratagene); Site 1:
 EcoRI; Site 2: XhoI; mRNA was purified from 4 x 10E5
 filariform larvae which had been isolated from infected
 humans. cDNA was constructed and, using adaptors, was
 cloned unidirectionally into the vector from the EcoRI
 site to the XhoI site. The library has an unamplified
 titer of 7 x 10E6 pfu/ml. The average insert size of the
 unamplified library is 975 bp (range, 500-1500)."

BASE COUNT
 ORIGIN

Query Match 11.6%; Score 36.6; DB 105; Length 446;
 Best Local Similarity 48.8%; Pred. No. 1.9;
 Matches 99; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 21 taatgttaacagccttcgacgaataatgctctagagatatagtgtaccatgccca 80
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 213 TAAATGTTGAACAGATCTCATCTCATATGATGAGATATTTGATATATGACT 154
 QY 81 aataaacaagagtagctctgtaaccaatgagtgactagctaccacacaaatg 140
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 153 ACTTGACATTATATTATCACGAGTGACGAATAATCTTTGATACGATATCATTTATATTC 94
 QY 141 cagtaactcagtgctcctcagcagcagctgataatgagagacagatgttttgatg 200
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 93 ATTATTTAATAATGATGATTCATTCATTAGGAAGTTGATTAATTTCTTGATATA 34
 QY 201 accagccagcagtggttgaat 223
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 33 ATTATTTGAACATTTGTTTCAT 11

RESULT 10
 AF046361/c 454 bp DNA GSS 21-APR-1998
 LOCUS Mus musculus clone OS115546, genomic survey sequence.
 DEFINITION AF046361
 ACCESSION AF046361.1 GI:3005232
 VERSION GSS.
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCE 1 (bases 1 to 454)
 Zambrowicz,B.P., Friedrich,G.A., Buxton,E.C., Lilleberg,S.L.,
 Person,C. and Sands,A.T.
 Disruption and sequence identification of 2,000 genes in mouse
 embryonic stem cells
 JOURNAL Nature 392 (6676), 608-611 (1998)
 MEDLINE 98219085
 REFERENCE 2 (bases 1 to 454)
 Zambrowicz,B.P., Friedrich,G.A., Buxton,E.C., Lilleberg,S.L.,
 Person,C. and Sands,A.T.
 Direct Submission
 JOURNAL Submitted (06-FEB-1998) OmniBank, Lexicon Genetics Incorporated,
 4000 Research Forest Drive, The Woodlands, TX 77381, USA
 FEATURES
 source
 1..454
 /organism="Mus musculus"
 /strain="129SV/Ev"
 /db_xref="taxon:10090"
 /clone_lib="OS115546"
 /cell_type="embryonic stem cell"
 /note="OmniBank Sequence Tag; exon trapped sequence"

BASE COUNT 155 a 104 c 94 g 70 t 31 others
 ORIGIN

Query Match 11.3%; Score 35.6; DB 190; Length 454;
 Best Local Similarity 46.0%; Pred. No. 3.8;
 Matches 98; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 98 ctgtatcccaatgagtgatgactgactttggccacaaatgacatgactgaatgccca 157
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 290 CTGCTGCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 231
 QY 158 ctggcactgcaactgtatgtagtgacagatgttttgatagatcagccagcagtg 217
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 230 CTGCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 171
 QY 218 ttaatgcaaaccttaacttataatgagtgatctcctcagggatgaagctcctggcc 277
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 170 TTGTTCTGTGCTACTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 111
 QY 278 ttcaggttttgcctgctgctgctgctgctgctgctgctgctgctgctgctgctg 310
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 110 CMTCTGTCNCNCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 78

RESULT 11
 CINS00F4U/c 928 bp DNA GSS 04-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC:
 DEFINITION BACR30H13 of RPCI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL070039.1 GI:4950182
 VERSION GSS.
 KEYWORDS fruit fly.
 SOURCE Drosophila melanogaster
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 928)
 Genoscope.
 TITLE Direct Submission
 AUTHORS Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage;
 JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 COMMENT - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila genome project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mamooser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
 P1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 source
 1..928
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="RPCI-98"
 /clone="BACR30H13"
 /note="end : 77"

BASE COUNT 268 a 242 c 192 g 184 t 42 others
 ORIGIN

Query Match 11.3%; Score 35.6; DB 190; Length 928;
 Best Local Similarity 53.6%; Pred. No. 4.7;
 Matches 59; Conservative 6; Mismatches 45; Indels 0; Gaps 0;

QY 143 gtactcagtgctcctcagcactgacatgagatgagatgttttgatagat 202

[illegible]

```

Oy 228 acctaaacttactataaagtggtgtcctcccaagggtgaagctccggcccttaagtttt 287
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 144 TGCTGTGCGATATAGAGTACGCTCTTGTGTGTCGTGCTGCTGCTGCTGCTTGTTC 85
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy 288 tgcgtcgtggtgctgcgcgtcgcagaagt 315
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84 TGCTGCTGCTGTTGCTGCTGCTGCTGCTT 57
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
CNS0115X
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - FRANCE (E-mail : segreff@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila genome project (EDGP) -
http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pbeloBAC11.

FEATURES
Source
1..849
location=Qualifiers
/organism="Drosophila melanogaster"
/plasmid="pbeloBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN05023"
/note="end : 77"

BASE COUNT 221 a 123 c 169 g 312 t 24 others

ORIGIN

Query Match 11.1%; Score 35.2; DB 190; Length 849;
Best Local Similarity 52.4%; Pred. No. 6;
Matches 76; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Oy 171 tgaatgagatgacagatgttttggatagatcacgcgcacagtgltgtaaatgcaaac 230
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 79 TGCTGCTGCTGTTGTTGCTGCTGCTGCTGTTGGTTCNGCTGTAGCTGTGAAGMAGCGC 138
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy 231 taacttactataatggtgtctccctccaaaggtgaagctcctgcctcgaagttttgc 290
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 139 GGTGTGTGACCGCTAATAGTTGGTGGCTGCTGCTGTTGATGTAGAGTTGCTGCTGC 198
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy 291 tgcgtgctgcgcgtcgcagaagt 315
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 199 TGCTGCTGCTGCTGCTGCTGTTGGT 223
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: March 6, 2001, 21:24:42
Job time: 22352 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:49:38 ; Search time 83.05 Seconds

(without alignments)
43.231 Million cell updates/sec

Title: US-09-196-161d-10

Perfect score: 560
Sequence: 1 GSAQGEANGNQPFANNAAR.....POGEAPGLQVFAAGAAAGCI 105

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_36:*

1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT:*

2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT:*

3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT:*

4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT:*

5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT:*

6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT:*

7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT:*

8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT:*

9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT:*

10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT:*

11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:*

12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT:*

13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT:*

14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT:*

15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:*

16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT:*

17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:*

18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT:*

19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:*

20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:*

21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	81	14.5	613	19 W73009	Cobra venom protea
2	81	14.5	621	19 W73013	Cobra venom mocran
3	75	13.4	1086	21 Y84111	Amino acid sequenc
4	75	13.4	1086	21 Y84116	Amino acid sequenc
5	75	13.4	1086	21 Y58835	Corn cellulose syn
6	72	12.9	320	19 W70220	Leishmania antigen
7	72	12.9	320	19 W70236	Leishmania antigen
8	72	12.9	410	20 W93577	Human APOB protein
9	72	12.9	468	19 W64483	Human DR4 protein.
10	72	12.9	468	20 Y31602	Human death recept
11	72	12.9	468	20 W93609	Human DR4 protein.
12	71	12.7	144	18 W50286	Human Fas antigen

13	71	12.7	159	18 W50288	Human Fas antigen
14	71	12.7	314	16 R76238	Fas-delta-TM. Hom
15	71	12.7	314	17 R99682	Human Fas soluble
16	71	12.7	335	13 R28084	Human cell surface
17	71	12.7	335	16 R28606	Human Fas protein.
18	71	12.7	335	17 R99681	Human Fas antigen.
19	71	12.7	335	17 R92528	hFas from plasmid
20	71	12.7	335	18 W50289	Human Fas antigen.
21	71	12.7	335	19 W49104	Fas protein. Mamm
22	71	12.7	376	18 W50287	Human Fas antigen
23	71	12.7	376	19 W60037	Antigenic peptide
24	71	12.7	600	16 R78610	Expression vector
25	71	12.7	669	19 W64484	Human TNFR1 protel
26	71	12.7	969	14 R41662	Paired basic amino
27	71	12.7	3110	16 R71730	Merostin major subu
28	71	12.7	3110	20 Y15460	Human laminin alph
29	70.5	12.6	685	18 W25719	Human alpha meltri
30	70.5	12.6	735	20 Y41695	Human PRO545 prote
31	70.5	12.6	1607	19 W50897	Mouse laminin G1 c
32	70	12.5	314	20 W98070	Soluble Fas recept
33	70	12.5	424	17 W04321	Ancylotoma secret
34	70	12.5	3084	19 W50891	Mouse laminin A ch
35	69.5	12.4	610	11 R05538	Endothelial-leukoc
36	69	12.3	1039	19 W73309	Cellulose synthase
37	69	12.3	1111	17 R91428	Kallinin/laminin 5
38	69	12.3	1193	17 R91427	Kallinin/laminin 5
39	69	12.3	3084	10 P94758	Sequence of mouse
40	68.5	12.2	495	20 Y59972	Human endometrium
41	68.5	12.2	610	11 R05494	Endothelial leukoc
42	68.5	12.2	610	11 R08116	Endothelial cell-1
43	68.5	12.2	610	18 W18839	E-selectin. Homo
44	68.5	12.2	610	19 W46733	Amino acid sequenc
45	68.5	12.2	610	21 Y59500	Human ELAM-1 prote

ALIGNMENTS

RESULT 1	
ID W73009	standard; protein; 613 AA.
XX AC W73009;	
XX AC	
DT 02-FEB-1999	(first entry)
XX DE	Cobra venom protease mocranagin NM-9.
XX KW	Mocranagin; snake venom; Mozambiquan spitting cobra; protease;
KW	Inflammation; myocardial infarction; thrombosis; infection;
KW	metastasis; therapy; NM-9.
XX OS	Naja mossambica mossambica.
XX XX	
XX FH	Key
FH	Location/Qualifiers
FT Peptide	1..23
FT	/label= sig_peptide
FT Protein	24..613
FT	/label= Mat_Protein
PN W09846771-A2.	
XX PD	
XX PD	22-OCT-1998.
XX PF	
PF	14-APR-1998; 98MO-US07998.
XX PR	
PR	18-FEB-1998; 98US-0026001.
PR	15-APR-1997; 97US-0843373.
XX PR	23-JAN-1998; 98US-0012637.
XX PA	
PA	(GEMV) GENETICS INST INC.
XX	
PI	Boodhoo A, Sako D, Seehra JS, Shaw G;

XX		MPI: 1998-568735/48.
DR	N-P8DB; V07897.	
XX		
PT	Isolated mocarhagin cobra venom protease, and nucleic acids encoding	
PR	it - used to develop products for treating e.g. myocardial	
PR	infarction, thrombosis, bacterial or viral infection, metastatic	
PR	conditions or inflammatory disorders	
PS	Claim 40; Page 45-48; 97pp; English.	
XX		
CC	This is the amino acid sequence of mocarhagin NMN-9, a highly	
CC	specific metalloprotease from the venom of the Mozambiquan	
CC	spitting cobra. The invention provides mocarhagin polypeptides	
CC	(see W73007-13) and polynucleotides (see V07895-901) encoding them,	
CC	as well as host cells and methods of producing the (especially	
CC	mature) polypeptides. Mocarhagin proteins are capable of cleaving	
CC	antionic polypeptide containing sulphated tyrosine residues,	
CC	p-selectin glycoprotein (GP) ligand-1 (PSGL-1) and GP1b-alpha	
CC	(claimed). They also inhibit neutrophil/HM60 binding, inhibit	
CC	platelet binding to von Willebrand factor, require Ca ²⁺ and Zn ²⁺	
CC	ions for activity and have activity inhibited by excess EDTA or	
CC	high concentrations of DFP (claimed).. They can be used to inhibit	
CC	selectin-mediated binding and to treat inflammatory disease	
CC	(claimed). In particular, they can be used to treat e.g. myocardial	
CC	infarction, vessel stenosis, thrombosis, bacterial or viral	
CC	infection, metastatic conditions, inflammatory disorders such as	
CC	arthritis, acute respiratory distress syndrome, asthma, emphysema,	
CC	delayed type hypersensitivity reaction, systemic lupus	
CC	erythematosus, thermal injury such as burns or frostbite,	
CC	autoimmune thyroiditis, experimental allergic encephalomyelitis,	
CC	multiple sclerosis, multiple organ injury syndrome secondary to	
CC	trauma, diabetes, Reynaud's syndrome, neutrophilic dermatosis	
CC	(Sweet's syndrome), inflammatory bowel disease, Grave's disease,	
CC	glomerulonephritis, gingivitis, periodontitis, haemolytic uremic	
CC	syndrome, ulcerative colitis, Crohn's disease, necrotising	
CC	enterocolitis, granulocyte transfusion associated syndrome,	
CC	cyclokinine-induced enterocolitis, granulocyte transfusion associated	
CC	syndrome, or cyclokinine-induced toxicity. Mocarhagin protein may	
CC	also be useful in organ transplantation, both to prepare organs for	
CC	transplantation and to quell organ transplant rejection, to treat	
CC	haemodialysis and leukopenia patients, or as an inhibitor of p-	
CC	- or E-selectin-mediated intercellular adhesion.	
XX		
SQ	Sequence 613 AA;	
XX		
Query Match	14.5%; Score 81; DB 19; Length 613;	
Best Local Similarity	28.1%; Pred. No. 1.1;	
Matches 27; Conservative 11; Mismatches 36; Indels 22; Gaps 5;		
OY	11 OPFAANMARAGICVBCQINRWGSGTINAG---DLATLATQCSTQCPTGTALDDGVTVDFD 66 : : : : : : : :	
Dd	443 qbhagcd-segcceckcfkfgagaecraakddcdldpelctcggaacp-----ctifq 492	
OY	67 RSAACVCVKRPFFXNGGSP-----QGAEAGLOV 95 : : : : :	
Dd	493 rnljpc-qneegyoyngkcplmtncglatlgpykvxv 527	
RESULT	2	
ID	W73013 standard; Protein; 621 AA.	
W73013		
XX	W73013;	
AC		
XX		
DT	02-FEB-1999 (first entry)	
XX		
XX	Cobra venom mocarhagin NMN-9ek.	
DE		
XX		
XX	Mocarhagin: snake venom; Mozambiquan spitting cobra; protease;	
KM	Inflammation; myocardial infarction; thrombosis; infection;	
KW	metastasis; therapy; NMN-9ek.	

XX	OS	Naja mossambica mossambica.	
XX	OS	Synthetic.	
XX	FH	Key	Location/Qualifiers
XX	FH	Peptide	1..196
XX	FT	Cleavage-site	/label= pro-peptide
XX	FT	Protein	192..196
XX	FT		/note= "enterokinase cleavage site"
XX	FT		197..621
XX	FT		/label= Mat_protein
XX	PN	MO9846771-A2.	
XX	PD	22-OCT-1998.	
XX	PF	14-APR-1998;	98WO-US07998.
XX	PR	18-FEB-1998;	98US-0026001.
XX	PR	15-APR-1997;	97US-0843373.
XX	PR	23-JAN-1998;	98US-0012637.
XX	PA	(GEMV) GENETICS INST INC.	
XX	PI	Boodhoo A, Sako D, Seehra JS, Shaw G;	
XX	DR	WPI: 1998-568735/48.	
XX	DR	N-PSDB: V07901.	
XX	PT	Isolated mocarhagin cobra venom protease, and nucleic acids encoding	
XX	PT	it - used to develop products for treating e.g. myocardial	
XX	PT	infarction, thrombosis, bacterial or viral infection, metastatic	
XX	PT	conditions or inflammatory disorders	
XX	PS	Claim 80; Page 65-68; 97pp; English.	
XX	PS	This is the amino acid sequence of a modified cobra venom mocarhagin	
XX	CC	protein, termed NNM-9ek, that includes an enterokinase cleavage site	
XX	CC	between the propeptide and mature peptide of mocarhagin (see also	
XX	CC	W73009). Introduction of the cleavage site may allow secretion of	
XX	CC	active mocarhagin from eukaryotic host cells. The invention	
XX	CC	provides mocarhagin polypeptides (see W73007-13) and polynucleotides	
XX	CC	(see V07895-901), as well as host cells and methods of producing	
XX	CC	(especially mature) polypeptides. Mocarhagin proteins are capable	
XX	CC	of cleaving anionic polypeptide containing sulphated tyrosine	
XX	CC	residues, P-selectin glycoprotein (Gp) 11gand-1 (PSG1-1) and	
XX	CC	GpIb-alpha (claimed). They also inhibit neutrophil/HL60 binding,	
XX	CC	inhibit platelet binding to von Willebrand Factor, require Ca2+ and	
XX	CC	zn2+ ions for activity and have activity inhibited by excess EDTA	
XX	CC	or high concentrations of DFP (claimed). They can be used for	
XX	CC	inhibiting selectin-mediated binding and for treating an	
XX	CC	inflammatory disease (claimed). In particular, they can be used	
XX	CC	for treating e.g. myocardial infarction, vessel restenosis,	
XX	CC	thrombosis, bacterial or viral infection, metastatic conditions,	
XX	CC	inflammatory disorders such as arthritis, acute respiratory distress	
XX	CC	syndrome, asthma, emphysema, delayed type hypersensitivity reaction,	
XX	CC	systemic lupus erythematosus, thermal injury e.g. burns or frostbite	
XX	CC	autoimmune thyroiditis, experimental allergic encephalomyelitis,	
XX	CC	multiple sclerosis, multiple organ injury syndrome secondary to	
XX	CC	trauma, diabetes, Reynaud's syndrome, neutrophilic dermatosis	
XX	CC	(Sweet's syndrome), inflammatory bowel disease, Grave's disease,	
XX	CC	glomerulonephritis), gingivitis, periodontitis, hemolytic uremic	
XX	CC	syndrome, ulcerative colitis, Crohn's disease, necrotising	
XX	CC	enterocolitis, granulocyte transfusion associated syndrome,	
XX	CC	cytokine-induced enterocolitis, granulocyte transfusion associated	
XX	CC	syndrome, or cytokine-induced toxicity. Mocarhagin protein may	
XX	CC	also be useful in organ transplantation, both to prepare organs for	
XX	CC	transplantation and to quell organ transplant rejection, to treat	
XX	CC	haemodialysis and leukopenia patients, or as an inhibitor of P-	
XX	CC	or E-selectin-mediated intercellular adhesion.	
XX	Sequence	621 AA;	

```

Query Match      14.5%: Score 81; DB 19; Length 621;
Best Local Similarity 28.1%; Pred. No. 1.1;
Matches    27; Conservative   11; Mismatches   36; Indels   22; Gaps     5

Oy       11 OPFAANMARGLICVCQOLIRVSGTGNAG----DLATLATQCSTOCPTGTALDGVDPED 66
          || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db        451 gphagcd-segcceckcfkfgagaectraakddclpelctcggaacp-----tdifrg 500

Oy         67 RSAACVKCKPNEYINGSP-----GGEAPGLGY 95
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db          501 rnglpc-gmegycykqkcplmtngcialsrpgvkv 535

RESULT      3
ID           Y84I11 standard; Protein: 1086 AA.
XX           Y84I11:
XX           Y84I11:
DT            03-JUL-2000 (first entry)

DE           Amino acid sequence of a maize cellulose synthase.
XX           Maize; cellulose synthase; stalk quality; strand; silage; cellulose;
XX           transgenic plant; plant breeding marker.
OS           Zea mays.
PN            WO200009706-A2.
PD            24-FEB-2000.
XX           PE
PE            16-AUG-1999; 99WO-US18760.
XX           PR
PR            17-AUG-1998; 98US-0096822.
PA            (PION-) PIONEER HI-BRED INT INC.
XX           PI
PI            Dhingra KS, Helentjaris TG, Bowen BA, Wang X;
XX           WPI: 2000-224343/19.
DR            N-PSDB; Z99503.

New genes which encode maize cellulose synthase polypeptides in plants
useful for modulating the expression of cellulose synthase in plants
and to produce transgenic plants expressing the novel protein -

Claim 15; Page 118-120; 119pp; English.

The present sequence represents a maize cellulose synthase polypeptide.
The cellulose synthase can be used for the improvement of stalk quality
for improved stand or silage. It also provides an increased concentration
of cellulose in the pericarp, hardening the kernel and improving its
handling ability. The sequences are used to produce transgenic plants
and seeds expressing the cellulose synthase. The polynucleotide is
used for modulating, preferably increasing, the level of the synthase
in a plant cell. The plants are preferably monocots. The polynucleotide
is also used as a probe or primer in the detection quantitation or
isolation of gene transcripts. The probes are useful in detecting
deficiencies in the level of mRNA in screenings for desired transgenic
plant, for detecting mutations in the gene, for monitoring upregulation
of expression or changes in enzyme activity in screening assays of
compounds, for detection of any number of allelic variants of the gene,
or for use as molecular markers in plant breeding programs. The
isolated nucleic acids of the present invention can also be used for
recombinant expression of their encoded polypeptides or for use as
immunogens in the preparation and/or screening of antibodies. The
proteins can be employed in assays for enzyme agonists or antagonists
of enzyme function or for use of immunogens or antigens to obtain
antibodies specifically immunoreactive with a protein.
```

```

Query Match Similarity      13.4%; Score 75; DB 21; Length 1086;
Best Local Similarity      27.4%; Pred No. 9;
Matches      31; Conservative      8; Mismatches      50; Indels      24; Gaps      4.

Db      43 gddvlgapggdfvvacneafvpcdcyeyrreggn-----cpqctlyrkllgcr 96
Qy      52 PGTGALDGCVTDFV-----RSAQCVKCKPFFYNGSGSPQGEAPGLQY 95
Db      97 vegdeedgdvddldnefnwdgdsqsvaesmljghmsygrgdgpnqapafq1 149

RESULT      4
ID      Y84116 standard; Protein: 1086 AA.
XX
AC      Y84116;
XX
DT      03-JUL-2000 (first entry)
XX
DE      Amino acid sequence of a maize cellulose synthase.
XX
KW      Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
KM      transgenic plant; plant breeding marker.
XX
OS      Zea mays.
XX
PN      WO200009706-A2.
XX
PD      24-FEB-2000.
XX
PF      16-AUG-1999; 99WO-US18760.
XX
PR      17-AUG-1998; 98US-0096822.
XX
PA      (PION-) PIONEER HI-BRED INT INC.
XX
PI      Dhuga KS, Helentjaris TG, Bowen BA, Wang X;
XX
DR      WPI: 2000-224343/19.
XX
N-PSDB: Z89518.
XX
PS      Claim 15; Page 157-160; 119pp; English.
XX
XX
XX      The present sequence represents a maize cellulose synthase polypeptide.
CC      The cellulose synthase can be used for the improvement of stalk quality
CC      for improved stand or silage. It also provides an increased concentration
CC      of cellulose in the pericarp, hardening the kernel and improving its
CC      handling ability. The sequences are used to produce transgenic plants
CC      and seeds expressing the cellulose synthase. The polynucleotide is
CC      used for modulating, preferably increasing, the level of the synthase
CC      in a plant cell. The plants are preferably monocots. The polynucleotide
CC      is also used as a probe or primer in the detection quantitation or
CC      isolation of gene transcripts. The probes are useful in detecting
CC      deficiencies in the level of mRNA in screenings for desired transgenic
CC      plants for detecting mutations in the gene, for monitoring upregulation
CC      of expression or changes in enzyme activity in screening assays of
CC      compounds, for detection of any number of allelic variants of the gene,
CC      or for use as molecular markers in plant breeding programs. The
CC      isolated nucleic acids of the present invention can also be used for
CC      recombinant expression of their encoded polypeptides or for use as
CC      immunogens in the preparation and/or screening of antibodies. The
CC      proteins can be employed in assays for enzyme agonists or antagonists
CC      of enzyme function or for use of immunogens or antigens to obtain
CC      antibodies specifically immunoreactive with a protein.
XX

```



```

RESULT 7
ID W70236 standard; Protein; 320 AA.
XX
AC W70236;
XX
DT 13-NOV-1998 (first entry)
XX
DE Leishmania antigen protein.
XX
KW Leishmania antigen; immune response; infection detection; therapy;
KW humoral response induction; cellular response induction; cancer;
KW Interleukin-12 production.
XX
OS Leishmania sp.
XX
PN W09835045-A2.
XX
PD 13-AUG-1998.
XX
PF 12-FEB-1998; 98WO-US03002.
XX
PR 27-AUG-1997; 97US-0920609.
PR 12-FEB-1997; 97US-0798841.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos Neto A, Dillon DC, Reed SG, Skeiky YA, Webb JR;
XX
DR WPI: 1998-447242/38.
XX
PT New immunogenic fragments of Leishmania antigens and related nucleic
PT acid, vectors and host cells - are useful for diagnosis, prevention
PT and treatment of leishmaniasis, also to induce production of
PT interleukin-12 generally
XX
PS Claim 56; Page 131-132; 194pp; English.
XX
CC This sequence represents a Leishmania antigen (LAg) of the
CC invention. Compositions and vaccines containing the protein are
CC used to generate a protective or therapeutic immune response against the
CC Leishmania species donavani, chagasi, infantum, major, amazonensis,
CC braziliensis, panamensis, tropica or guayanaensis. They can also be used
CC to detect infection (in a skin test). The compositions induce a humoral
CC and/or cellular response, specifically of Th1 type, particularly
CC including induction of interleukin-12 (IL-12) production. They may thus
CC be used more generally to treat any condition (e.g. bacterial, viral or
CC protozoal infection, or cancer) which responds to IL-12.
XX
SQ Sequence 320 AA;

Query Match 12.9%; Score 72; DB 19; Length 320;
Best Local Similarity 31.2%; Pred. NO. 4.7;
Matches 25; Conservative 9; Mismatches 24; Indels 22; Gaps 7;

QY 9 GNOFFAANNARGICVPCOINRVSGTGNAGDLATLATQCTGTALDDGV-----T 62
   |::| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 67 gsepl--n-----mc-pc-----tdpncascpsdagtc-tgcangyldvgacvrcqep 112

QY 63 DVF--DRSAOCVCKPKNFY 80
   | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 113 ncfscsdankctgcapy 132

RESULT 8
ID W93577 standard; Protein; 410 AA.
XX
AC W93577;
XX
DT 18-JUN-1999 (first entry)
XX

```

```

DE Human APOB protein.
XX
KW Tumour necrosis factor receptor; signal transducer molecule; TNF; APO4;
KW developmental abnormality; gestational abnormality; prostate cancer;
KW APO6; APO8; APO9; TNRL-1; TNRL-3; diagnosis; treatment; therapy; disease;
KW cytoplasmic domain; immunogen; antibody preparation; breast carcinoma;
KW apoptosis; human; APOB; APO-related protein.
XX
OS Homo sapiens.
XX
PN W09911791-A2.
XX
PD 11-MAR-1999.
XX
PF 04-SEP-1998; 98WO-US18393.
XX
PR 05-SEP-1997; 97US-0924634.
XX
PA (UNIV ) UNIV WASHINGTON.
XX
PI Chaudhary PM;
XX
DR WPI: 1999-205191/17.
DR N-PSDB; X23411.
XX
PT New Tumor Necrosis Factor family receptor polypeptides and ligands -
PT useful for diagnosis and treatment of prostate cancer and
PT developmental or gestational abnormalities
XX
PS Example 1; Fig 3; 156pp; English.
XX
CC This invention describes isolated Tumor Necrosis Factor (TNF) family
CC receptor polypeptides: APO4, APO6, APO8 and APO9 or their active
CC fragments, and isolated TNF related ligands 1 and 3 (TNRL1 and TNRL3) or
CC their active fragments. APO4 is useful for diagnosing prostate cancer
CC by determining levels of APO4 in an individual. Prostate cancer can also
CC be treated using APO4 selective binding agents linked to a therapeutic
CC moiety. APO4 polypeptides are also useful for identifying selective
CC binding agents, useful in diagnosis/treatment of disease, by binding of
CC agents to the polypeptide/active fragment which is extracellular, or
CC expressed on the cell surface. The binding is preferably performed in
CC vivo. APO4 polypeptides/active fragments are also useful for screening
CC for agonists and antagonists by binding and observing the change in APO4
CC activity. Effective pharmacological agents useful in diagnosis or
CC treatment of disease are also identified using APO4 polypeptides/active
CC fragments and APO4 signal transducer molecules that specifically interact
CC with a cytoplasmic domain of APO4 and detecting a change in level of APO4
CC activity. The method is performed in vivo or in vitro. APO polypeptides
CC are all useful as immunogens for preparing antibodies. APO4 is also
CC useful for diagnosis/treatment of developmental or gestational
CC abnormalities. APO8 was transfected to human breast carcinoma cell line
CC MCF-7, and induced apoptosis.
XX
SQ Sequence 410 AA;

Query Match 12.9%; Score 72; DB 20; Length 410;
Best Local Similarity 25.3%; Pred. NO. 6.2;
Matches 24; Conservative 8; Mismatches 37; Indels 26; Gaps 4;

QY 5 GEANGNOFFAANNARGI-----CVPCOINRVSGTGNAGDLATLATQCTSTCPT 53
   | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 77 gshserpgacnrtcgygynasnlfacipctacksdeerspctttrntac--qckp 134

QY 54 GTALDDGVTVDFDRSAOCVCKPKNFYINGSSPOG 88
   | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 135 glfrn-----dnsaemckrc-----slgcpvg 156

RESULT 9
ID W64483 standard; Protein; 468 AA.
XX
AC W64483
XX
DT
XX

```

```

AC W64483;
XX 20-OCT-1998 (first entry)
XX
XX
DE Human DR4 protein.
XX
XX Death domain containing receptor 4; DR4; apoptosis; cancer; inflammation;
KW agonist; tumour necrosis factor; TNF; ligand; autoimmune disease;
KW infection; graft rejection; antagonist; inhibitor; diagnostic.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX Peptide 1..23
XX FT /label= signal
XX FT Protein 24..468
XX FT /label= DR4
XX FT Domain 24..238
XX FT /label= extracellular_domain
XX FT Domain 239..264
XX FT /label= transmembrane_domain
XX FT Domain 265..468
XX FT /label= intracellular_domain
XX FT Domain 379..422
XX FT /label= death_domain

WO9832856-A1.
XX 30-JUL-1998.
XX
XX 27-JAN-1998; 98WO-US01464.
XX
XX 05-FEB-1997; 97US-0037829.
XX 28-JAN-1997; 97US-0035722.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (UNMI) UNIV MICHIGAN.
XX
XX Dixit VM, Gentz RL, NI J, Pan JG, Rosen CA:
XX WPI: 1998-427952/36.
XX DR N-PSDB; V49527.
XX
XX Nucleic acid encoding human death domain-containing receptor 4 -
XX PT useful for therapeutic modulation of apoptosis, in e.g. cancer and
XX PT autoimmune diseases
XX PS
XX Claim 1a: Fig 1; 92pp; English.
XX
XX This sequence represents a human death domain containing receptor 4, DR4.
XX DR4 agonists are used to increase apoptosis induced by tumour necrosis
XX factor (TNF)-family ligands, e.g. in cases of cancer, autoimmune disease,
XX viral or other infections, inflammation, graft vs. host disease, acute or
XX chronic graft rejection. Antagonists of DR4 are used to inhibit such
XX apoptosis, e.g. in cases of acquired immune deficiency syndrome,
XX neurodegenerative disease, myelodysplastic syndrome, ischaemic injury,
XX toxin-induced liver damage, septic shock, cachexia and anorexia, also a
XX wide range of inflammatory conditions. DR4 of fragments of the protein
XX are used diagnostically, e.g. to detect mutant forms of DR4 (possibly
XX associated with disease), for isolating the DR4 gene or related sequences
XX and for chromosomal mapping.
XX
XX Sequence 468 AA;
XX
XX
XX Query Match 12.9%; Score 72; DB 19; Length 468;
XX Best Local Similarity 25.3%; Pred. NO. 7.2;
XX Matches 24; Conservative 8; Mismatches 37; Indels 26; Gaps 4;
XX
XX 5 GEANGNOPFAANNAARGI-----CVPCOINRVSGSTNAGDLATATOCSTOCP 53
XX :| | | | | :| | | | | :| | | | |
XX 135 gshserpgaanrctegvygynasnlfacipctacksdeerspctttrntac--gckp 192
XX

```

```

OY 54 GTALDDGVTVDFDRSAQCCKPNEYNGSGPOG 88
XX :| | | | | :| | | | | :| | | | |
XX Db 193 gtfnn-----dhsaemcrkc-----stgcp 214
XX
XX
XX RESULT 10
XX Y31602
XX ID Y31602 standard; Protein; 468 AA.
XX
XX AC Y31602;
XX
XX 09-NOV-1999 (first entry)
XX
XX Human death receptor-4.
XX
XX TNF receptor; tumour necrosis factor receptor; cell surface receptor;
XX antibody; Apo-2 ligand; TRAIL ligand; apoptosis; DR4; cancer.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT Domain 1..218
XX FT /label= extracellular

WO9937684-A1.
XX 29-JUL-1999.
XX
XX 25-JAN-1999; 99WO-US01437.
XX
XX 26-JAN-1998; 98US-0072481.
XX
XX (GETH-) GENENTECH INC.
XX
XX Chuntharapal A, Kim KJ:
XX WPI: 1999-469117/39.
XX DR N-PSDB; Z08960.
XX
XX New antibodies to death receptor-4, used for modulating activities
XX PT associated with Apo-2 ligand, particularly apoptosis, useful for
XX PT treating diseases and pathological conditions, e.g. cancer
XX
XX Disclosure: Fig 1; 21pp; English.
XX
XX The present sequence is a human death receptor-4 (DR4) protein, a
XX CC member of the tumour necrosis factor receptor family which is involved in
XX CC apoptosis induction. DR4 is also thought to be a TRAIL and Apo-2
XX CC ligand. The protein is used to produce antibodies (monoclonal or
XX CC chimeric) that specifically bind to DR4. The DR4 antibodies may be
XX CC agonistic, antagonistic or blocking antibodies. The DR4 antibodies are
XX CC capable of modulating biological activities associated with Apo-2 ligand,
XX CC in particular, apoptosis, and thus are useful in the treatment of various
XX CC diseases and pathological conditions, including cancer. The antibodies
XX CC can also be used for disease detection and diagnosis.
XX
XX Sequence 468 AA;
XX
XX
XX Query Match 12.9%; Score 72; DB 20; Length 468;
XX Best Local Similarity 25.3%; Pred. NO. 7.2;
XX Matches 24; Conservative 8; Mismatches 37; Indels 26; Gaps 4;
XX
XX 5 GEANGNOPFAANNAARGI-----CVPCOINRVSGSTNAGDLATATOCSTOCP 53
XX :| | | | | :| | | | | :| | | | |
XX Db 135 gshserpgaanrctegvygynasnlfacipctacksdeerspctttrntac--gckp 192
XX
XX 54 GTALDDGVTVDFDRSAQCCKPNEYNGSGPOG 88
XX :| | | | | :| | | | | :| | | | |
XX Db 193 gtfnn-----dhsaemcrkc-----stgcp 214
XX
XX RESULT 11

```

W93609
ID W93609 standard; Protein; 468 AA.
XX
AC W93609;
XX
DE 18-JUN-1999 (first entry)
XX
DE Human DR4 protein.
XX
KW Killer protein; adriamycin inducible; human; chromosome 8p21; diagnosis;
KW p53-inducible; apoptosis-mediating activity; treatment; animal model;
KW neoplastic disease; DR4.
XX
OS Homo sapiens.
XX
PN M09902653-A1.
XX
PD 21-JAN-1999.
XX
PF 10-JUL-1998; 98WO-US14495.
XX
PR 11-MAR-1998; 98US-0077661.
PR 11-JUL-1997; 97US-0052305.
PR 04-AUG-1997; 97US-0054710.
PR 30-SEP-1997; 97US-0060473.
PR 11-MAR-1998; 98US-0077526.
PR 11-MAR-1998; 98US-0077628.
XX
PA (UYPE-) UNIV PENNSYLVANIA.
XX
PI El-Deiry WS;
XX
DR WPI: 1999-120857/10.
XX
PT A new nucleic acid encodes a p53-induced protein (Killer) - which
PT induces apoptosis and is useful in the diagnosis and treatment of
PT neoplastic diseases
XX
PS Disclosure: Page 45; 65pp; English.
XX
CC This invention describes a novel human adriamycin-inducible killer
CC protein located on chromosome 8p21, which also has p53-inducible,
CC apoptosis-mediating activity and comprises an amino-terminal
CC extracellular receptor, transmembrane and death domains. The nucleic
CC acid molecule which encodes the protein, it's encoded signal
CC transduction protein and antibodies of the invention are useful in the
CC diagnosis and treatment of neoplastic diseases. The invention is also
CC useful for the production of animal model systems.
XX
SQ Sequence 468 AA;
SO

Query Match 12.9%; Score 72; DB 20; Length 468;
Best Local Similarity 25.3%; Pred. NO. 7.2;
Matches 24; Conservative 8; Mismatches 37; Indels 26; Gaps 4;

OY 5 GEANMGPFANNAARGI-----CVPQINRVSGTNGADLTATLTCSTGCP 53
| | | | | : | | | | | : | | | | |
Db 135 gshserpgacnctegvgynasnlfacipctackdeerspcttlnlac--gckp 192
| | | | | : | | | | | : | | | | |
OY 54 GTALDDCVTDVDRSAOCVCKKPNFYNGSSPOG 88
| | | | | : | | | | | : | | | | |
Db 193 gtfrr-----dnasemcrkc-----stgcprrg 214
| | | | | : | | | | | : | | | | |

RESULT 12
W50286
ID W50286 standard; Protein; 144 AA.
XX
AC W50286;
XX
DT 16-JUL-1998 (first entry)
XX

DE Human Fas antigen derivative.
XX
KW Human; Fas antigen; derivative; apoptosis regulation; gene therapy;
KW treatment; diabetes; arthritis; lupus; hepatitis; influenza; HIV;
KW apoptosis modulation.
XX
OS Homo sapiens.
XX
FH Key
FH Peptide 1..16
FT /label= sig-peptide
FT Peptide 17..144
FT /label= mat-peptide
XX
PN M09742319-A1.
XX
PD 13-NOV-1997.
XX
PF 01-MAY-1997; 97WO-JP01502.
XX
PR 02-MAY-1996; *96JP-0135760.
XX
PA (MOCH) MOCHIDA PHARM CO LTD.
PA (OSAB-) OSAKA BIOSCIENCE INST.
XX
PI Nagata S, Nakamura N;
XX
DR WPI: 1997-558981/51.
DR N-PSDB: V07003.
XX
PT Fas antigen derivative containing modified extracellular region -
PT has low antigenicity, promotes apoptosis and is useful in treatment
PT of viral and other diseases
XX
PS Disclosure: Fig 3; 102pp; Japanese.
XX
CC The present sequence is a Fas antigen derivative, which
CC contain a Fas antigen extracellular region lacking one or more
CC amino acid residues in the region from the amino-terminal to (but
CC excluding) the 1st cysteine residue (preferably at least 29
CC residues are deleted)
CC The derivative is an effective regulator of apoptosis and can be
CC used (either by administration of the polypeptide, or by the use
CC of the coding DNA in gene therapy) to treat a range of diseases,
CC e.g. diabetes, arthritis, lupus and in particular viral diseases
CC such as hepatitis, influenza and HIV, by modulating apoptosis of
CC virus-infected cells.
XX
SQ Sequence 144 AA;
SO

Query Match 12.7%; Score 71; DB 18; Length 144;
Best Local Similarity 28.6%; Pred. NO. 2.5;
Matches 22; Conservative 10; Mismatches 27; Indels 18; Gaps 4;

OY 8 NGNPGFANNAARGICVPCQINRVSGTNGADLTATLTC--STGCPGTALDDGVTDVF 65
| | | | | : | | | | | : | | | | |
Db 47 ngdep-----dvcpcq-----egkeycdkahnfskrrrrclcdghglevein--c 90
| | | | | : | | | | | : | | | | |
OY 66 DRSAOCVCKKPNFYNN 82
| | | | | : | | | | | : | | | | |
Db 91 trtgnkrcrkpnffcn 107
| | | | | : | | | | | : | | | | |

RESULT 13
W50288
ID W50288 standard; Protein; 159 AA.
XX
AC W50288;
XX
DT 16-JUL-1998 (first entry)
XX
DE Human Fas antigen derivative/IgG1 hinge fusion.

FT	/note= "soluble Fas del1 antigen"
FT	17..168
FT	/label= Extracellular_domain
FT	/note= "the 5 C-terminal residues of the
FT	Fas antigen extracellular domain are
FT	deleted in Fas del1"
FT	Domain
FT	169..314
FT	/label= Cytoplasmic_domain
FT	Peptide
FT	164..173
FT	/note= "preferred peptide from breakpoint region
FT	(claim 4, page 132)"
FT	Peptide
FT	164..174
FT	/note= "preferred peptide from breakpoint region"
FT	161..171
FT	/note= "preferred peptide from breakpoint region"
XX	
PX	W09620206-A1.
PD	
PD	04-JUL-1996.
XX	
PE	22-DEC-1995; 95WO-US17083.
PR	23-DEC-1994; 94US-0371263.
XX	
PA	(UABR-) UAB RES FOUND.
PI	
PI	Cheng J, Liu C, Mountz JD, Zhou T;
XX	
DR	WPI. 1996-321796/32.
N-	N-PSDB; T34527.
PT	
PT	Natural, soluble form of Fas antigen secreted by human cells is
PT	result of alternative mRNA processing - used to diagnose
PT	Fas-associated disease, e.g. systemic lupus erythematosus
XX	
PS	Claim 4; Page 114-16; 152pp; English.
XX	
CC	A natural, soluble Fas antigen variant (R99682), designated Fas
CC	dell, and other Fas variants (R99683-85) are derived by alternative
CC	splicing of Fas gene transcripts. A cDNA clone (T34527) coding for
CC	the variant was obtd. from human peripheral blood mononuclear cells.
CC	The Fas dell variant lacks the transmembrane domain of insoluble
CC	Fas antigen (R99681). Recombinant dell variant, or fragments of
CC	it, can be expressed in prokaryotic or eukaryotic (e.g. COS) cells.
CC	Detection of increased levels of soluble forms of Fas antigen can
CC	be used to diagnose autoimmune diseases, esp. systemic lupus
CC	erythematosus and angioimmunoblastic lymphadenopathy.
XX	
SO	Sequence 314 AA;
Query Match	12.7%; Score 71; DB 17; Length 314;
Best Local Similarity	28.6%; Pred. No. 5.9;
Matches 22; Conservative 10; Mismatches 27; Indels 18; Gaps	
OY	8 NGNDFEANNARIGICVPCQINRVGSGTNAGDLATLATQC--STGCCPGTALDDGVTVF 65 :: : : :
Db	76 ngdep-----dcvpcq-----egkyekdkahtsfskccrrcricdeghgyleveIn--c 119 ::: ::
OY	66 DRSAAQCIVKCKPNEYNN 82
Db	120 trtgnrkckrcpkntfnc 136

```
Search completed: March  6, 2001, 12:49:39
Job time: 100 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:52:51 ; Search time 116.78 Seconds
(without alignments)
20.073 Million cell updates/sec

Title: US-09-196-161D-5
Perfect score: 106
Sequence: 1 CPTGTALDDGVTDVDFDRSAA 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

```
Searched:      374700 seqs, 117207915 residues
Total number of hits satisfying chosen parameters:  3747000
```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```

1:  SPREMEL.15:*
2:  SP_Archaea:*
3:  SP_Bacteria:*
4:  SP_Fungi:*
5:  SP_Human:*
6:  SP_Invertebrate:*
7:  SP_Mammal:*
8:  SP_mhc:*
9:  SP_Organelle:*
10: SP_Phage:*
11: SP_Plant:*
12: SP_Rodent:*
13: SP_Virus:*
14: SP_Unclassified:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	106	100.0	394	5	027208	Q27208 ichthyophth
2	106	100.0	442	5	09XG2	09XG2 ichthyophth
3	49	46.2	546	5	09N912	09N912 trypanosoma
4	46	43.4	362	2	P72142	P72142 pseudomonas
5	46	43.4	463	10	032145	032145 arabidopsis
6	46	43.4	466	1	05S591	05S591 pyrococcus
7	45	42.5	2747	2	09L800	09L800 aeromonas s
8	44	41.5	111	3	09P608	09P608 neurospora
9	44	41.5	398	10	09SXI9	09SXI9 zeia mayas (m
10	44	41.5	424	5	076744	076744 nectaror ame
11	44	41.5	424	5	09XZ41	09XZ41 ancylostoma
12	44	41.5	425	5	077153	077153 ancylostoma
13	43	40.6	175	3	000300	000300 atellomyces
14	43	40.6	175	3	09P439	09P439 atellomyces
15	43	40.6	175	3	09P438	09P438 atellomyces
16	43	40.6	175	3	09P437	09P437 atellomyces
17	43	40.6	175	3	09P436	09P436 atellomyces
18	43	40.6	175	3	09P435	09P435 atellomyces
19	43	40.6	175	3	09P434	09P434 atellomyces

20	43	40.6	175	3	09p433	ajelomyces
21	43	40.6	175	3	09p341	ajelomyces
22	43	40.6	360	3	09p8b3	agarcus bi
23	43	40.6	475	12	082009	agarcus bi
24	43	40.6	778	13	09tbc4	human papi
25	43	40.6	2180	5	001768	xenopus la
26	43	40.6	2183	11	088783	caenorhabdi
27	42.5	40.1	255	5	09xux7	mus musculi
28	42.5	40.1	372	5	045332	caenorhabdi
29	42.5	40.1	527	5	09xux3	caenorhabdi
30	42	39.6	87	10	03lpe1	caenorhabdi
31	42	39.6	183	2	p94595	arabidopsi
32	42	39.6	256	3	p79000	saccharomyc
33	42	39.6	305	2	045818	chloroflexu
34	42	39.6	461	10	049610	arabidopsi
35	42	39.6	468	4	09tuc32	homo sapien
36	42	39.6	767	5	076977	strongyloce
37	42	39.6	1104	4	060460	homo sapien
38	42	39.6	1138	10	092w00	arabidopsi
39	42	39.6	1283	4	095451	homo sapien
40	42	39.6	1373	4	075372	homo sapien
41	42	39.6	2806	2	09kxa6	escherichia
42	42	39.6	2806	9	09xjml	bacterioph
43	42	39.6	2806	9	09n1k9	bacterioph
44	41.5	39.2	2806	3	059934	ophiostoma
45	41	38.7	173	1	027632	methanobact

ALIGNMENTS

RESULT	1			
027208				
AC	027208	PRELIMINARY;	PRF;	394 AA.
DT	01-NOV-1996	(TREMBLrel. 01, Created)		
DT	01-NOV-1998	(TREMBLrel. 08, last sequence update)		
DT	01-JUN-2000	(TREMBLrel. 14, last annotation update)		
DE	IMMOBILIZATION ANTIGEN PRECURSOR (FRAGMENT).			
OS	Ichthyophthirius multifiliis.			
OC	Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida.			
CC	Opisthokontina; Ichthyophthirius.			
OX	NCBI_TaxID=5932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=GEORGIA;			
RX	MEDLINE=92335298; PubMed=1631132;			
RA	Clark T.G., McGraw R.A., Dickerson H.W.;			
RT	"Developmental expression of surface antigen genes in the parasitic			
RT	ciliate Ichthyophthirius multifiliis.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 89:6363-6367(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=GEORGIA;			
RX	MEDLINE=93020590; PubMed=1383510;			
RA	Lin T.L., Dickerson H.W.;			
RT	"Purification and partial characterization of immobilization antigens			
RT	from Ichthyophthirius multifiliis.";			
RL	J. Protozool. 39:457-463(1992).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=GEORGIA;			
RA	Clark T.;			
RL	Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=GEORGIA;			
RA	Clark T.;			
RL	Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: M92907; AAC36158.1; ..			
KW	Signal.			
FT	NON_TER	1	1	
FT	SIGNAL	<1	1	POTENTIAL.

```

SQ SEQUENCECEE 394 AA; 39495 MW; 3013C2B2BEFDB682 CRC64;

Query Match          100.0%; Score 106; DB 5; Length 394;
Best Local Similarity 100.0%; Pred. No. 3.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 CPTGTALDDGVTDFDRSAA 20
        |||
Db       86 CPTGTALDDGVTDFDRSAA 105

RESULT 2
O9XZG2 PRELIMINARY; PRT: 442 AA.
AC O9XZG2;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE IMMOBILIZATION ANTIGEN PRECURSOR.
DE IAG48.
OS Ichthyophthirius multifiliis.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
CC Ophryotrocha; Ichthyophthirius.
OX NCBI_TaxID=5932;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=G1;
RC MEDLINE=99196987; PubMed=10095108;
RX Clark T.G., Lin T.L., Jackwood D.A., Sherrill J., Lin Y.,
RA Dickerson H.W.;
RT "The gene for an abundant parasite coat protein predicts tandemly
   repetitive metal binding domains.";
RL Gene 229:91-100(1999).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=G1;
RC Gaertig J., Gao Y., Tischgarten T., Clark T.G., Dickerson H.W.;
RA "Surface display of a parasite antigen in the ciliate Tetrahymena
   thermophila.";
RT Nat. Biotechnol. 0:0-0(1999);
DR EMBL; AF140273; AAD31283.1; -.
KW Signal.
FT SIGNAL.
FT CHAIN 1 20 POTENTIAL.
FT CHAIN 21 442 IMMOBILIZATION ANTIGEN.
SQ SEQUENCE 442 AA; 45025 MW; 52658F3F65D27AFA CRC64;

Query Match          100.0%; Score 106; DB 5; Length 442;
Best Local Similarity 100.0%; Pred. No. 4.3e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 CPTGTALDDGVTDFDRSAA 20
        |||
Db       106 CPTGTALDDGVTDFDRSAA 125

RESULT 3
O9N912 PRELIMINARY; PRT: 546 AA.
AC O9N912;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE HYPOTHEtical 59.2 KDA PROTEIN.
DE CHR1.59.
OS Trypanosoma brucei.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5691;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=TREN927M;
RA Hall N., Bowman S., Quail M., Ivens A.C., Kay M.P., Bray-Allen S.,
```

```

RA Lennard N.J., Clark L.N., Harris B.R., Melville S., Lawson D.,
RA Gerrard C., Rajadream M.A., Barrell B.G.;
RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL359782; CAB95371.1; -.
KW Hypothetical protein.
SQ SEQUENCE 546 AA; 59236 MW; 5526205781B8289 CRC64;

OY 4 GTALDDGVTDVDFDRSA 19
   |:::|::|::|
Db 531 GVPIDDGLTDAFRRA 546

RESULT 4
P72142 PRELIMINARY; PRT; 362 AA.
ID P72142
AC P72142;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE WPPI.
GN WPPI.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCHI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAOI;
RC MEDLINE=97093969; PubMed=8939432;
RX Burrows L.L., Charter D.F., Lam J.S.;
RT "Molecular characterization of the Pseudomonas aeruginosa serotype O5
RT (PAOI) B-band lipopolysaccharide gene cluster.";
RL MOL. MICROBIOL. 22:481-495(1996).
DR EMBL; U05061; AAC45863.1; -.
SO SEQUENCE 362 AA; 39745 MW; 251076C00FEF0844 CRC64;

Query Match 46.2%; Score 49; DB 5; Length 546;
Best Local Similarity 56.2%; Pred. NO. 9.2;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 CPTGTALDDGVTDVDFDRSA 20
   ||| |::|::|::|
Db 146 CPTRYAIDNLKNEGFERKAA 165

RESULT 5
O22145 PRELIMINARY; PRT; 463 AA.
ID O22145
AC O22145;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE PUTATIVE SIALOGLYCOPROTEASE.
GN F4123.22.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCHI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RC Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
RA Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D., Somerville C.R.,
RA Venter J.C.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC0002367; AAB82636.1; -.

```

DR INTERPRO: IPR000905; -.
 DR PfAM: PF00814; Peptidase_M22; 1.
 DR PRINTS: PR00789; OSIALOPTASE.
 DR PRODOM: PD002367; -. 1.
 KW Protease.
 SO SEQUENCE 463 AA; 51267 MW; E50377F24E8A59D CRC64;

Query Match 43.4%; Score 46; DB 10; Length 463;
 Best Local Similarity 43.8%; Pred. No. 23;
 Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 4 GTALDDGVTVDFRSA 19
 |||:::|::|::|
 Db 235 GTTVDDAIGEAFFDKTA 250

RESULT 6
 ID 059591 PRELIMINARY; PRT; 466 AA.
 AC 059591;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE HYPOTHEICAL 50.5 KDA PROTEIN PH1928.
 GN PH1928.
 OS Pyrococcus horikoshii.
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
 OX NCBI_TaxID=53953;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OT3;
 RX MEDLINE=98344137; PubMed=9679194;
 RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y., Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y., Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohnuku Y., Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K., Masuchi Y., Shizuya H., Kikuchi H.;
 RA "Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
 RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
 RL DNA Res. 5:55-76(1998).
 DR EMBL: AP000007; BAA1055.1; -.
 KW Hypothetical protein.
 SO SEQUENCE 466 AA; 50519 MW; 6B7975D40FC8DE17 CRC64;

Query Match 43.4%; Score 46; DB 1; Length 466;
 Best Local Similarity 66.7%; Pred. No. 24;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 6 ALDDGVTVDFPR 17
 |||::|::|::|
 Db 399 ALDEGIVDMDR 410

RESULT 7
 ID 091800 PRELIMINARY; PRT; 2747 AA.
 AC 091800;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE RTX PROTEIN.
 GN ASX.
 OS Aeromonas salmonicida.
 OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
 OC Aeromonas.
 OX NCBI_TaxID=645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33658;
 DR Braun M., Frey J., Kuhnert P.;

RT "280 kDa RTX protein of Aeromonas.";
 RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF218037; AAF27914.1; -.
 SO SEQUENCE 2747 AA; 280202 MW; 208FE380E44A5F37 CRC64;

Query Match 42.5%; Score 45; DB 2; Length 2747;
 Best Local Similarity 75.0%; Pred. No. 2.5e+02;
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 TGTALDDGVTV 14
 |||||::|::|
 Db 354 TGTATDDGSDV 365

RESULT 8
 ID 09P608 PRELIMINARY; PRT; 111 AA.
 AC 09P608;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE HYPOTHEICAL 11.9 KDA PROTEIN.
 GN B208.90.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Schulte U., Aign V., Hohelsel J., Brandt P., Fartmann B., Holland R., Nyakatura G., Mewes H.W., Mannhaupt G.;
 RA Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA German Neurospora genome project;
 RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: ALJ35930; CAB91362.1; -.
 KW Hypothetical protein.
 SO SEQUENCE 111 AA; 11863 MW; 00A9EDB3EB058AA5 CRC64;

Query Match 41.5%; Score 44; DB 3; Length 111;
 Best Local Similarity 61.5%; Pred. No. 9.9;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 CPTGTALDDGVTD 13
 ||::|::|::|
 Db 28 CPSSAALDDGTYD 40

RESULT 9
 ID 09SX19 PRELIMINARY; PRT; 398 AA.
 AC 09SX19;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE CYSTEINE PROTEASE MITR1.
 GN MITR1.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 CC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
 OX NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=CALLUS;
 RX MEDLINE=99320873; PubMed=10394950;
 RA Pechan T., Jiang B., Steckler D., Ye L., Lin L., Luthe D.S., Williams W.P.;
 RA "Characterization of three distinct cDNA clones encoding cysteine proteinases from maize (Zea mays L.) callus.";
 RT Plant Mol. Biol. 40:111-119(1999).
 RL EMBL: AF019145; AAB70820.2; -.

DR HSSP; P00785; 2ACT.
DR INTERPRO: IPR000169; -
DR INTERPRO: IPR000668; -
DR PFAM: PF00112; Peptidase_C1; 1.
DR PRINTS: PR00705; PAPAIN.
DR PROSITE: PS00138; THIOI_PROTEASE_CYS; 1.
DR PROSITE: PS00639; THIOI_PROTEASE_HIS; 1.
DR PROSITE: PS00640; THIOI_PROTEASE_ASN; 1.
DR PROTEASE.
KW
SEQUENCE 398 AA; 42624 MW; 6BF2CE89CB82E80E CRC64;

Query Match 41.5%; Score 44; DB 10; Length 398;
Best Local Similarity 69.2%; Pred. No. 42;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 PGTGALDDGVTDV 14
DB 310 PGGTSLDHGVAV 322

RESULT 10

ID 076744 PRELIMINARY; PRT; 424 AA.
AC 076744;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE ANCYLOSTOMA SECRETED PROTEIN 1 PRECURSOR.
GN ASP1.
OS Necator americanus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Strongyilitida;
OC Ancylostomatoidae; Ancylostomatidae; Bunostominae; Necator.
OX NCBI_TaxID=51031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SHANGHAI;
RX MEDLINE=96215086; PubMed=8636085;
RA Hawdon J.M., Jones B.F., Hoffman D.R., Hotez P.J.;
RT "Cloning and characterization of Ancylostoma-secreted protein. A novel
RT protein associated with the transition to parasitism by infective
RT hookworm larvae.";
RL J. Biol. Chem. 271:6672-6678(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SHANGHAI;
RA Zhan B., Hawdon J.M., Shan Q., Ren H., Qiang H., Hu W., Xiao S.-H.,
RA Li T., Gong X., Feng Z., Hotez P.J.;
RT "Ancylostoma Secreted Protein (ASP) Homologs in Human Hookworms.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR HSSP; P04284; ICPE.
DR INTERPRO: IPR001283; -
DR PFAM: PF00188; SCP; 1.
DR PRINTS: PR00837; V5TPXLIKE.
KW Signal.
FT CHAIN 1 18 POTENTIAL.
FT SIGNAL 19 424 ANCYLOSTOMA SECRETED PROTEIN 1.
SEQUENCE 424 AA; 45742 MW; BF1EB2F95F9B4A9F CRC64;

Query Match 41.5%; Score 44; DB 5; Length 424;
Best Local Similarity 46.7%; Pred. No. 45;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 1 CPTGTALDDGVTDV 15
DB 227 CPSNTGMTDSVDRVF 241

RESULT 11

OYXZ41
ID 09XZ41 PRELIMINARY; PRT; 424 AA.

AC 09XZ41;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE ANCYLOSTOMA-SECRETED PROTEIN 1 PRECURSOR.
OS Ancylostoma caninum (Dog hookworm).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Strongyilitida;
OC Ancylostomatoidae; Ancylostomatidae; Ancylostominae; Ancylostoma.
OX NCBI_TaxID=29170;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SHANGHAI;
RA Zhan B., Shan Q., Hawdon J.M.;
RT "Variation between Asp-1 molecules from Ancylostoma caninum in China
RT and the US.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF132291; AADJ1839.1; -
DR HSSP; P04284; ICPE.
DR INTERPRO: IPR001283; -
DR INTERPRO: IPR002413; -
DR PFAM: PF00188; SCP; 1.
DR PRINTS: PR00837; V5TPXLIKE.
DR PRINTS: PR00838; V5ALLERGEN.
KW Signal.
FT CHAIN 1 18 POTENTIAL.
FT SIGNAL 19 424 ANCYLOSTOMA-SECRETED PROTEIN 1.
SEQUENCE 424 AA; 45761 MW; 8409CDDF8AECDD248E CRC64;

Query Match 41.5%; Score 44; DB 5; Length 424;
Best Local Similarity 46.7%; Pred. No. 45;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 1 CPTGTALDDGVTDV 15
DB 227 CPSNTGMTDSVDRVF 241

RESULT 12

ID 077153 PRELIMINARY; PRT; 425 AA.
AC 077153;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE ANCYLOSTOMA-SECRETED PROTEIN 1 PRECURSOR.
GN ASP1.
OS Ancylostoma duodenale.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Strongyilitida;
OC Ancylostomatoidae; Ancylostomatidae; Ancylostominae; Ancylostoma.
OX NCBI_TaxID=51022;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SHANGHAI;
RX MEDLINE=96215086; PubMed=8636085;
RA Hawdon J.M., Jones B.F., Hoffman D.R., Hotez P.J.;
RT "Cloning and characterization of Ancylostoma-secreted protein. A novel
RT protein associated with the transition to parasitism by infective
RT hookworm larvae.";
RL J. Biol. Chem. 271:6672-6678(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SHANGHAI;
RA Zhan B., Hawdon J.M., Shan Q., Ren H., Qiang H., Hu W., Xiao S.-H.,
RA Li T., Gong X., Feng Z., Hotez P.J.;
RT "Ancylostoma-secreted protein (ASP) homologs in human hookworms.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF077402; AAD13339.1; -
DR INTERPRO: IPR001283; -
DR PFAM: PF00188; SCP; 1.
DR PRINTS: PR00837; V5TPXLIKE.
KW Signal.
FT SIGNAL 1 19 POTENTIAL.

FT CHAIN 20 425 , ANCILOSTOMA-SECRETED PROTEIN 1.
SQ SEQUENCE 425 AA; 43821 MW; 1F6B9D7E62EEDDA8 CRC64;

Query Match 41.5%; Score 44; DB 5; Length 425;
Best Local Similarity 46.7%; Pred. No. 45;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 1 CPTGTALDDGVTVDF 15
||:|:|:|:|:|
Db 228 CPSTNGMTDSVDRDF 242

RESULT 13
ID Q00300 PRELIMINARY; PRT; 175 AA.

AC Q00300:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE TEMPERATURE-DEPENDENT PROTEIN Bysl (BLASTOMYCES YEAST PHASE-SPECIFIC PROTEIN 1).
GN Bysl.
OS Ajellomyces dermatitidis (Blastomyces dermatitidis).
OC Eukaryota; Fungi; Ascomycota; Onygenales; Onygenaceae; Ajellomyces.
OX NCBI_TaxID=5039;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-T58;
RX MEDLINE=94245364; PubMed=8188377;
RA Burg E.F., Smith L.H.;
RT "Cloning and characterization of bysl, a temperature-dependent CDNA specific to the yeast phase of the pathogenic dimorphic fungus Blastomyces dermatitidis.";
RL Infect. Immun. 62:2521-2528(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-LE;
RA Bono J.L., Jaber B., Fisher M.A., Abuodeh R.O., O'Leary-Jepson E., Scalapone G.M., Smith L.H. Jr.;
RT "Genetic diversity and transcriptional analysis of the bysl gene from Blastomyces dermatitidis.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: U09020; AAA50955.1; -
DR EMBL: AF277089; AAF6483.1; -
SQ SEQUENCE 175 AA; 18558 MW; E8AE072A4008EF06 CRC64;

Query Match 40.6%; Score 43; DB 3; Length 175;
Best Local Similarity 77.8%; Pred. No. 24;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 CPTGTALDD 9
||:|:|:|:|
Db 163 CPGTSLDD 171

RESULT 14

ID Q9P439 PRELIMINARY; PRT; 175 AA.

AC Q9P439:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE BLASTOMYCES YEAST PHASE-SPECIFIC PROTEIN 1.
GN Bysl.
OS Ajellomyces dermatitidis (Blastomyces dermatitidis).
OC Eukaryota; Fungi; Ascomycota; Onygenales; Onygenaceae; Ajellomyces.
OX NCBI_TaxID=5039;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-2;
RA Bono J.L., Jaber B., Fisher M.A., Abuodeh R.O., O'Leary-Jepson E.,

RA Scalapone G.M., Smith L.H. Jr.;
RT "Genetic diversity and transcriptional analysis of the bysl gene from Blastomyces dermatitidis.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF277079; AAF6473.1; -
SQ SEQUENCE 175 AA; 18606 MW; 8288BDD98F38AA12 CRC64;

Query Match 40.6%; Score 43; DB 3; Length 175;
Best Local Similarity 77.8%; Pred. No. 24;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 CPTGTALDD 9
||:|:|:|:|
Db 163 CPGTSLDD 171

RESULT 15

ID Q9P438 PRELIMINARY; PRT; 175 AA.

AC Q9P438:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE BLASTOMYCES YEAST PHASE-SPECIFIC PROTEIN 1.
GN Bysl.
OS Ajellomyces dermatitidis (Blastomyces dermatitidis).
OC Eukaryota; Fungi; Ascomycota; Onygenales; Onygenaceae; Ajellomyces.
OX NCBI_TaxID=5039;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-103;
RA Bono J.L., Jaber B., Fisher M.A., Abuodeh R.O., O'Leary-Jepson E., Scalapone G.M., Smith L.H. Jr.;
RT "Genetic diversity and transcriptional analysis of the bysl gene from Blastomyces dermatitidis.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF277080; AAF6474.1; -
SQ SEQUENCE 175 AA; 18468 MW; 498BC73B612104AA CRC64;

Query Match 40.6%; Score 43; DB 3; Length 175;
Best Local Similarity 77.8%; Pred. No. 24;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 CPTGTALDD 9
||:|:|:|:|
Db 163 CPGTSLDD 171

Search completed: March 6, 2001, 12:52:54
Job time: 294 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:49:34 ; Search time 83.05 Seconds
(without alignments)
6.176 Million cell updates/sec

Title: US-09-196-161d-6
Perfect score: 94
Sequence: 1 CVCKRNFYNGSP 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues
Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_36:*

- 1: /SIDSI/gcgdata/geneseq/geneseq/AA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseq/AA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseq/AA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseq/AA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseq/AA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseq/AA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseq/AA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseq/AA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseq/AA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseq/AA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseq/AA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseq/AA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseq/AA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseq/AA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseq/AA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseq/AA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseq/AA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseq/AA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseq/AA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseq/AA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseq/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	46.8	58	13	R27401 Human neutrophil e
2	44	46.8	58	17	R99159 Genetically engine
3	43	45.7	23	12	R15280 Anti-thrombin fusi
4	43	45.7	37	21	R91614 Human secreted pro
5	43	45.7	286	21	R91613 Human secreted pro
6	43	45.7	1607	19	W50897 Mouse laminin G1 c
7	43	45.7	1609	19	W50898 Human laminin G1 c
8	42	44.7	20	19	W83054 Human Fas peptide
9	42	44.7	20	21	W90912 Human Fas extracel
10	42	44.7	20	21	Y68518 Human Fas peptide
11	42	44.7	58	13	R27396 Human neutrophil e
12	42	44.7	58	13	R27397 Human neutrophil e

13	42	44.7	58	13	R27400 Human neutrophil e
14	42	44.7	58	13	R27402 Human neutrophil e
15	42	44.7	58	13	R27394 Human neutrophil e
16	42	44.7	58	13	R27395 Human neutrophil e
17	42	44.7	58	17	R99158 Genetically engine
18	42	44.7	58	17	R99160 Genetically engine
19	42	44.7	58	17	R99162 Genetically engine
20	42	44.7	58	17	R99163 Genetically engine
21	42	44.7	58	17	R99164 Genetically engine
22	42	44.7	124	18	W25933 New protease inhib
23	42	44.7	125	20	W85556 Heavy chain V regi
24	42	44.7	144	18	W50286 Human Fas antigen
25	42	44.7	144	18	W25938 SUC2-Ep7-d21-RPDR
26	42	44.7	159	18	W50288 Human Fas antigen
27	42	44.7	235	18	W20672 H. pylori surface
28	42	44.7	242	18	W20472 H. pylori transmem
29	42	44.7	292	20	W85607 Secreted protein c
30	42	44.7	314	16	R76238 Fas-delta-TM. Hom
31	42	44.7	314	17	R99682 Human Fas soluble
32	42	44.7	314	20	W98070 Soluble Fas recept
33	42	44.7	335	13	R28084 Human cell surface
34	42	44.7	335	16	R78606 Human Fas protein.
35	42	44.7	335	17	R99681 Human Fas antigen.
36	42	44.7	335	17	R92528 hFas from plasmid
37	42	44.7	335	18	W50289 Human Fas antigen.
38	42	44.7	335	19	W49104 Fas protein. Mamm
39	42	44.7	376	18	W50287 Human Fas antigen
40	42	44.7	376	19	W60037 Antigenic peptide
41	42	44.7	516	20	Y07735 Human breast-speci
42	42	44.7	600	16	R78610 Expression vector
43	42	44.7	600	17	R92526 Fas antigen #1. S
44	42	44.7	669	19	W64484 Human TNFR1 protei
45	42	44.7	3075	19	W50892 Human laminin A ch

ALIGNMENTS

RESULT 1		ALIGNMENTS	
R27401	ID	R27401 standard; peptide; 58 AA.	
XX	XX		
AC	R27401;		
XX	XX		
DT	20-MAY-1998 (first entry)		
XX	XX		
DE	Human neutrophil elastase inhibitor MOTT26A.		
XX	XX		
KW	Kunitz Domain serine protease inhibitor; mutant;		
KW	Bovine pancreatic trypsin inhibitor.		
XX	XX		
OS	Synthetic.		
XX	XX		
FH	Key	Location/Qualifiers	
FT	region	1..4	/note="mutated from ITI-D1 (KEDS) -> (RPDF)"
FT	region	15..19	/note="mutated from ITI-D1 (MGMTS) -> (VAMPF)"
FT	region	26	/note="mutated from ITI-D1 (MGMTS) -> (VAMPF)"
FT	region	31..34	/note="mutated from ITI-D1 (Y) -> (A)"
FT	region	/note="mutated from ITI-D1 (FTFO) -> (CTFV)"	
XX	XX		
PN	W09215605-A.		
XX	XX		
PD	17-SEP-1992.		
XX	XX		
PF	28-FEB-1992;	92WO-US01501.	
XX	XX		
PR	01-MAR-1991;	91US-0664989.	
PR	17-JUN-1991;	91US-0715834.	
XX	XX		
PA	(PROT-) PROTEIN ENG CORP.		

XX Guterman SK, Kent RB, Ladner RC, Ley AC, Markland W, Roberts BL,
 XX WPI; 1992-331666/40.
 XX
 PT New peptide inhibitors of elastase or cathepsin G - are e.g.
 PT mutants of Kunitz Domain serine protease inhibitors, useful for
 PT treating and preventing conditions caused by excessive neutrophil
 CC elastase or cathepsin G
 XX
 PS Example; Page 103; 126pp; English.
 XX
 CC The sequence is that of a mutant of inter-alpha trypsin inhibitor-D1
 CC (ITI-D1) MUTT26A which is an inhibitor with very strong (K_{0.5} 10⁻¹¹)
 CC specific binding activity for human neutrophil elastase (hNE). It can
 CC be used for the treatment or prophylaxis of a condition caused by
 CC excessive hNE activity, e.g. inflammation, emphysema, cystic fibrosis,
 CC adult respiratory distress syndrome or rheumatoid arthritis. It may
 CC also be used to purify hNE. See also R27373-R27403 and R27443.
 CC
 SQ Sequence 58 AA:

Query Match 46.8%; Score 44; DB 13; Length 58;
 Best Local Similarity 50.0%; Pred. No. 5.2;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 1 CVKCKPNEYNGS 14
 11 | : : : : : |
 DB 14 cvamfpryfingas 27

RESULT 2
 R99159
 ID R99159 standard; protein; 58 AA.
 AC R99159;
 XX
 DT 12-FEB-1997 (first entry)
 XX
 DE Genetically engineered aprolinin-like Kunitz domain (MUTT26A).
 XX
 KW Aprolinin; Kunitz domain; human neutrophil elastase; hNE;
 KW connective tissue; alpha 1 protease inhibitor; API; neutrophil;
 KW alpha1 antitrypsin; respiratory disorder; cystic fibrosis;
 KW smokers emphysema.
 XX
 OS Synthetic.
 XX
 PN WO9620278-A2.
 XX
 PD 04-JUL-1996.
 XX
 PF 15-DEC-1995; 95WO-US16349.
 XX
 PR 16-DEC-1994; 94US-0358160.
 XX
 PA (PROT-) PROTEIN ENG CORP.
 XX
 PI Guterman SK, Kent RB, Ladner RC, Ley AC, Markland W;
 PI Roberts BL;
 XX
 DR WPI; 1996-321851/32.
 XX
 PT New engineered inhibitors of human neutrophil elastase - contg.
 PT aprolinin-like Kunitz domain for treating, e.g. cystic fibrosis or
 PT other respiratory disorders
 XX
 PS Claim 1; Page 48; 105pp; English.
 XX
 CC Genetically engineered human derived Kunitz domains can be used to
 CC inhibit human neutrophil elastase, an enzyme involved in the
 CC elimination of pathogens and the restructuring of connective tissue.

CC In cases of reduction of the circulating alpha-1-protease inhibitor
 CC (API or alpha1 antitrypsin), or the inactivation of API by oxidation
 CC (smokers emphysema), extensive destruction of the lung tissue may
 CC result from uncontrolled elastolytic activity of human neutrophil
 CC elastase. Other respiratory disorders such as cystic fibrosis are
 CC thought to be caused by human neutrophil elastase release by
 CC neutrophils. The genetically engineered human derived Kunitz
 CC domains can be used to treat such respiratory disorders. See
 CC R99146-R99211.
 XX
 SQ Sequence 58 AA:

Query Match 46.8%; Score 44; DB 17; Length 58;
 Best Local Similarity 50.0%; Pred. No. 5.2;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 1 CVKCKPNEYNGS 14
 11 | : : : : : |
 DB 14 cvamfpryfingas 27

RESULT 3
 R15280
 ID R15280 standard; Protein; 23 AA.
 AC R15280;
 XX
 DT 17-FEB-1992 (first entry)
 XX
 DE Anti-thrombic fusion peptide (8).
 XX
 KW Molecular recognition unit; MRU; CDR3; PAC-1 antibody;
 KW platelet fibrinogen receptor; effector domain; metallothionein.
 XX
 OS Synthetic.
 XX
 FT Key Location/Qualifiers
 FT 7..23
 FT Domain /label= MRU_domain
 FT 1..6
 FT Domain /label= effector_domain
 XX
 FT WO9117173-A.
 XX
 PD 14-NOV-1991.
 XX
 PF 01-MAY-1991; 91WO-US03116.
 XX
 PR 07-MAY-1990; 90US-0519702.
 XX
 PA (CYTO-) CYTOGEN CORP.
 XX
 PI Rodwell JD, McKearn TJ, Alvarez VL, Radcliffe RD;
 PI WPI; 1991-353714/48.
 XX
 DR New conjugates of molecular recognition units - having first
 DR binding domain for desired target site and second effector
 PT domain, useful in medicine, agriculture, waste management etc.
 XX
 PS Disclosure; Page 46; 74pp; English.
 XX
 CC Using the amino acid sequence of the CDR3 region of PAC-1 antibody
 CC as a starting point, a novel family of conjugates of MRUs was prepd.
 CC (see R15273-88). The conjugates comprise: an MRU domain which is the
 CC amino acid sequence encoded by CDR3 of the PAC-1 antibody or a portion
 CC of it, pref. engineered to have enhanced affinity, compared with the
 CC MRU obtained from the native CDR3, for activated platelet fibrinogen
 CC receptor and an effector domain which is the carboxyl terminal end
 CC of the amino acid sequence of metallothionein.
 CC The amino or carboxy terminus of the fusion peptide can be blocked.
 CC See also R15373-97.

XX Sequence 23 AA:
SQ

Query Match 45.7%; Score 43; DB 12; Length 23;
Best Local Similarity 40.0%; Pred. No. 2.8;
Matches 6; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 1 CVCKRPNFYNGSP 15
1 1 1 1 1 1 1
Db 2 ctccapsyrrgdgap 16

RESULT 4
Y91614
ID Y91614 standard; Protein; 37 AA.
XX
XX Y91614;
XX
XX 29-JUN-2000 (first entry)
XX
XX Human secreted protein sequence encoded by gene 17 SEQ ID NO:287.
DE
XX Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW antiHIV; antiinflammatory; neutropenic; neuroprotective; antiallergic;
KW osteopathic; antiarthritic; antibacterial; antidiabetic; antiasthma;
KW antiparasitic; cardiant; gene therapy; cancer; neurological disorder;
KW immune disease; inflammation; blood disorder; tumour.
XX
XX Homo sapiens.
OS
XX WO200006698-A1.
PN
XX 10-FEB-2000.
PD
XX 29-JUL-1999; 99WO-US17130.
PF
XX 30-JUL-1998; 98US-0094657.
PR 05-AUG-1998; 98US-0095486.
PR 06-AUG-1998; 98US-0095454.
PR 06-AUG-1998; 98US-0095455.
PR 12-AUG-1998; 98US-0096319.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Komatsoulis GA, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;
PI Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;
PI Soppet DR, Endress GA, Ebner R, Olsen HS, Mucenski M;
XX
XX WPI: 2000-195282/17.
DR
XX New isolated human genes and the secreted polypeptides they encode,
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PT
XX
XX Disclosure: Page 48; 634pp; English.
PS
XX The polynucleotide sequences given in A26346 to A26458 encode the human
CC secreted proteins given in Y91451 to Y91691. The human secreted proteins
CC can have activities based on the tissues and cells they are expressed in.
CC Examples of the activities are: cytostatic; immunosuppressive; antiHIV;
CC antiinflammatory; neutropenic; neuroprotective; antiallergic; osteopathic;
CC antiarthritic; antibacterial; antidiabetic; antiasthma; antiparasitic;
CC and cardiant. The polynucleotides and their corresponding secreted
CC proteins are useful for preventing, treating or ameliorating medical
CC conditions, e.g. by protein or gene therapy. Also pathological conditions
CC can be diagnosed by determining the amount of the proteins in a sample or
CC by determining the presence of mutations in the polynucleotides. Specific
CC uses are described for each of the polynucleotides, based on which
CC tissues they are most highly expressed in and include developing
CC products for the diagnosis or treatment of cancer, tumours,
CC neurodegenerative disorders, developmental abnormalities and foetal
CC deficiencies, blood disorders, diseases of the immune system,

CC autoimmune diseases, hepatic and renal disease, inflammation,
CC allergies, Alzheimer's and behavioural disorders, schizophrenia,
CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
CC cardiovascular disorders, reproductive disorders, gastrointestinal
CC disorders, respiratory disorders and metabolic disorders. The proteins
CC or polynucleotides can also be used as food additives or preservatives.
CC The proteins are also useful for identifying their binding partners.
CC A26337 to A26345 and Y91450 are sequences used in the exemplification of
CC the present invention.
XX
XX SQ Sequence 37 AA:
XX

Query Match 45.7%; Score 43; DB 21; Length 37;
Best Local Similarity 66.7%; Pred. No. 4.6;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 CVCKRPNFY 9
1 1 1 1 1 1 1
Db 13 cqrckpgfy 21

RESULT 5
Y91613
ID Y91613 standard; Protein; 296 AA.
XX
XX Y91613;
XX
XX 29-JUN-2000 (first entry)
DT
XX Human secreted protein sequence encoded by gene 17 SEQ ID NO:286.
DE
XX
XX Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW antiHIV; antiinflammatory; neutropenic; neuroprotective; antiallergic;
KW osteopathic; antiarthritic; antibacterial; antidiabetic; antiasthma;
KW antiparasitic; cardiant; gene therapy; cancer; neurological disorder;
KW immune disease; inflammation; blood disorder; tumour.
XX
XX Homo sapiens.
OS
XX WO200006698-A1.
PN
XX 10-FEB-2000.
PD
XX 29-JUL-1999; 99WO-US17130.
PF
XX 30-JUL-1998; 98US-0094657.
PR 05-AUG-1998; 98US-0095486.
PR 06-AUG-1998; 98US-0095454.
PR 06-AUG-1998; 98US-0095455.
PR 12-AUG-1998; 98US-0096319.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Komatsoulis GA, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;
PI Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;
PI Soppet DR, Endress GA, Ebner R, Olsen HS, Mucenski M;
XX
XX WPI: 2000-195282/17.
DR
XX New isolated human genes and the secreted polypeptides they encode,
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PT
XX
XX Disclosure: Page 47; 634pp; English.
PS
XX The polynucleotide sequences given in A26346 to A26458 encode the human
CC secreted proteins given in Y91451 to Y91691. The human secreted proteins
CC can have activities based on the tissues and cells they are expressed in.
CC Examples of the activities are: cytostatic; immunosuppressive; antiHIV;
CC antiinflammatory; neutropenic; neuroprotective; antiallergic; osteopathic;
CC antiarthritic; antibacterial; antidiabetic; antiasthma; antiparasitic;

CC and cardiant. The polynucleotides and their corresponding secreted
 CC proteins are useful for preventing, treating or ameliorating medical
 CC conditions, e.g. by protein or gene therapy. Also pathological conditions
 CC can be diagnosed by determining the amount of the proteins in a sample or
 CC by determining the presence of mutations in the polynucleotides. Specific
 CC uses are described for each of the polynucleotides, based on which
 CC tissues they are most highly expressed in, and include developing
 CC products for the diagnosis or treatment of cancer, tumours,
 CC neurodegenerative disorders, developmental abnormalities and foetal
 CC deficiencies, blood disorders, diseases of the immune system,
 CC autoimmune diseases, hepatic and renal disease, inflammation,
 CC allergies, Alzheimer's and behavioural disorders, schizophrenia,
 CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
 CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
 CC cardiovascular disorders, reproductive disorders, gastrointestinal
 CC disorders, respiratory disorders and metabolic disorders. The proteins
 CC or polynucleotides can also be used as food additives or preservatives.
 CC The proteins are also useful for identifying their binding partners.
 CC A26337 to A26345 and Y14450 are sequences used in the exemplification of
 CC the present invention.

CC
 XX Sequence 296 AA;

Query Match 45.7%; Score 43; DB 21; Length 296;
 Best Local Similarity 66.7%; Pred. No. 39;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 CVCKPNFY 9
 | : | | | | |
 Db 75 cqrckpgfy 83

RESULT 6
 ID W50897 standard; Protein; 1607 AA.
 XX

AC W50897;

DT 07-DEC-1998 (first entry)

XX Mouse laminin G1 chain.

XX Laminin; mouse; beta-amyloid; amyloidosis; Alzheimer's disease;
 KW Down's syndrome; hereditary cerebral haemorrhage; inflammation;
 KW malignancy; Familial Mediterranean Fever; multiple myeloma;
 KW type II diabetes; prion disease; Creutzfeldt-Jacob disease; CJD;
 KW Gerstmann-Strausler syndrome; kuru; scrapie; haemodialysis;
 KW carpal tunnel syndrome; senile cardiac amyloid polynuropathy;
 KW Familial Amyloidotic Polynuropathy; thyroid carcinoma; diagnosis;
 KW therapy.

XX Mus sp.

XX W09815179-A1.

XX 16-APR-1998.

XX 08-OCT-1997; 97WO-US18145.

XX 08-OCT-1996; 96US-0027981.

XX (UNIW) UNIV WASHINGTON.

XX Castillo G, Snow AD;

XX WPI; 1998-240534/21.

PT Use of laminin and fragments - for developing products for use in
 PT the diagnosis and treatment of amyloid disease, e.g. Alzheimer's
 PT disease or CJD

PS Claim 15; Page 102-105; 132pp; English.

XX This is the amino acid sequence of the mouse laminin G1 chain. The
 CC primary object of the invention is to use laminin, laminin-derived
 CC protein fragments and/or laminin-derived polypeptides as potent
 CC inhibitors of amyloid formation, deposition, accumulation and/or
 CC persistence in Alzheimer's disease and other amyloidoses. The
 CC laminin products (see W50888-98) may include mouse or human laminin
 CC A or A1 chain, laminin B1 or B2 chain, laminin A2 chain (merotin),
 CC laminin G1 chain, the globular repeats of the laminin A1 chain and
 CC the beta-amyloid binding domain of the laminin A chain. A claimed
 CC method for treating an amyloid disease comprises administering a
 CC polypeptide having a conformational similarity to a fragment of a
 CC laminin protein. A method for diagnosing an amyloid disease
 CC involves determining levels of laminin in a sample. Production
 CC of laminin or its fourth globular repeat in vivo provides a method
 CC for in vivo inhibition of beta-amyloid amyloidosis. The products
 CC and methods can be used for the diagnosis, prognosis, monitoring
 CC and treatment of amyloidoses such as Alzheimer's disease, Down's
 CC syndrome and hereditary cerebral haemorrhage with amyloidosis of
 CC the Dutch type (where the specific amyloid is the beta-amyloid
 CC protein), the amyloidosis associated with chronic inflammation,
 CC various forms of malignancy and Familial Mediterranean Fever (FA
 CC amyloid or inflammation-associated amyloidosis), the amyloidosis
 CC associated with multiple myeloma and other B-cell abnormalities
 CC (AL amyloid), the amyloidosis associated with type II diabetes
 CC (amylin or islet amyloid), the amyloidosis associated with prion
 CC diseases including Creutzfeldt-Jacob disease, Gerstmann-Strausler
 CC syndrome, kuru and animal scrapie (PrP amyloid), the amyloidosis
 CC associated with long-term haemodialysis and carpal tunnel syndrome
 CC (beta 2-microglobulin amyloid), the amyloidosis associated with
 CC senile cardiac amyloid and Familial Amyloidotic Polynuropathy
 CC (prealbumin or transthyretin amyloid), and the amyloidosis
 CC associated with endocrine tumours such as medullary carcinoma of
 CC the thyroid (variant of procalcitonin).

SO Sequence 1607 AA;

Query Match 45.7%; Score 43; DB 19; Length 1607;
 Best Local Similarity 40.0%; Pred. No. 2.2e+02;
 Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 1 CVCKPNFYNGSP 15
 | : | : | : | : |
 Db 1010 cdqceenyfynrswp 1024

RESULT 7
 ID W50898 standard; Protein; 1609 AA.

XX W50898;

DT 07-DEC-1998 (first entry)

XX Human laminin G1 chain.

XX Laminin; human; beta-amyloid; amyloidosis; Alzheimer's disease;
 KW Down's syndrome; hereditary cerebral haemorrhage; inflammation;
 KW malignancy; Familial Mediterranean Fever; multiple myeloma;
 KW type II diabetes; prion disease; Creutzfeldt-Jacob disease; CJD;
 KW Gerstmann-Strausler syndrome; kuru; scrapie; haemodialysis;
 KW carpal tunnel syndrome; senile cardiac amyloid polynuropathy;
 KW Familial Amyloidotic Polynuropathy; thyroid carcinoma; diagnosis;
 KW therapy.

XX Homo sapiens.

XX W09815179-A1.

XX 16-APR-1998.

XX 08-OCT-1997; 97WO-US18145.

XX 08-OCT-1996; 96US-0027981.
 XX (UNIT) UNIV WASHINGTON.
 XX
 XX Castillo G, Snow AD.
 XX WPI: 1998-240534/21.
 DR
 XX Use of laminin and fragments - for developing products for use in
 PT the diagnosis and treatment of amyloid disease, e.g. Alzheimer's
 PT disease or CTD
 XX
 XX Claim 15; Page 106-109; 132pp; English.
 PS
 XX This is the amino acid sequence of the human laminin G1 chain. The
 CC primary object of the invention is to use laminin, laminin-derived
 CC protein fragments and/or laminin-derived polypeptides as potent
 CC inhibitors of amyloid formation, deposition, accumulation and/or
 CC persistence in Alzheimer's disease and other amyloidoses. The
 CC laminin products (see W50888-98) may include mouse or human laminin
 CC A or A1 chain, laminin B1 or B2 chain, laminin A2 chain (merosin),
 CC laminin G1 chain, the globular repeats of the laminin A1 chain and
 CC the beta-amyloid binding domain of the laminin A chain. A claimed
 CC method for treating an amyloid disease comprises administering a
 CC polypeptide having a conformational similarity to a fragment of a
 CC laminin protein. A method for diagnosing an amyloid disease
 CC involves determining levels of laminin in a sample. Production
 CC of laminin or its fourth globular repeat in vivo provides a method
 CC for in vivo inhibition of beta-amyloid amyloidosis. The products
 CC and methods can be used for the diagnosis, prognosis, monitoring
 CC and treatment of amyloidoses such as Alzheimer's disease, Down's
 CC syndrome and hereditary cerebral haemorrhage with amyloidosis of
 CC the Dutch type (where the specific amyloid is the beta-amyloid
 CC protein), the amyloidosis associated with chronic inflammation,
 CC various forms of malignancy and Familial Mediterranean Fever (AA
 CC amyloid or inflammation-associated amyloidosis), the amyloidosis
 CC associated with multiple myeloma and other B-cell abnormalities
 CC (AL amyloid), the amyloidosis associated with type II diabetes
 CC (amylin or islet amyloid), the amyloidosis associated with prion
 CC diseases including Creutzfeldt-Jacob disease, Gerstmann-Strausler
 CC syndrome, kuru and animal scrapie (PrP amyloid), the amyloidosis
 CC associated with long-term haemodialysis and carpal tunnel syndrome
 CC (beta 2-microglobulin amyloid), the amyloidosis associated with
 CC senile cardiac amyloid and Familial Amyloidotic Polyneuropathy
 CC (prealbumin or transthyretin amyloid), and the amyloidosis
 CC associated with endocrine tumours such as medullary carcinoma of
 CC the thyroid (variant of procalcitonin).
 CC
 XX
 XX Sequence 1609 AA;
 SQ
 Query Match 45.7%; Score 43; DB 19; Length 1609;
 Best Local Similarity 40.0%; Pred. NO. 2.2e+02;
 Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
 QY 1 CVKCRPNFYNGSP 15
 I : I : I : I : I : I :
 Db 1012 cdgceenyfynswp 1026
 RESULT 8
 ID W83054 standard; Peptide; 20 AA.
 XX W83054:
 AC
 XX 15-MAR-1999 (first entry)
 DT
 XX Human Fas peptide p12.
 DE
 XX Human Fas peptide p12.
 KW Fas; epitope; HFE7A; monoclonal antibody; humanised antibody;
 KW human; apoptosis; HFE7A; autoimmune disease; Hashimoto's disease;

KW systemic lupus erythematosus; graft versus host disease;
 KW Sjogren syndrome; pernicious anaemia; Addison's disease;
 KW scleroderma; Goodpasture syndrome; Crohn's disease; sterility;
 KW rheumatoid arthritis; autoimmune haemolytic anaemia;
 KW myasthenia gravis; multiple sclerosis; Basedow's disease;
 KW thrombopenia purpura; insulin-dependent diabetes; allergy;
 KW atopy; arteriosclerosis; myocarditis; cardiomyopathy;
 KW glomerular nephritis; hypoplastic anaemia; hepatitis; AIDS;
 KW transplant rejection; therapy.
 OS
 XX Synthetic.
 OS Homo sapiens.
 XX
 XX AU9859701-A.
 PN
 XX
 XX 08-OCT-1998.
 PD
 XX
 XX 30-MAR-1998; 98AU-0059701.
 PF
 XX
 XX 08-OCT-1997; 97JP-0276064.
 PR
 XX 01-APR-1997; 97JP-0082953.
 PR
 XX 25-JUN-1997; 97JP-0169088.
 XX
 XX (SANY) SANKYO CO LTD.
 PA
 XX
 XX Akio S, Hidayuki H, Hiroko Y, Jun O, Kimihisa I;
 PI Masahiko O, Nobufusa S, Shin Y, Tohru T;
 PI WPI: 1998-543440/47.
 DR
 XX
 XX New antibodies and proteins bind conserved epitope of Fas antigen -
 PT used to evaluate drugs in animal models and to treat Fas-associated
 PT diseases e.g. autoimmune disease, allergy, atopy, arteriosclerosis,
 PT myocarditis, hepatitis and AIDS
 XX
 XX Reference Example 6; Page 87; 292pp; English.
 PS
 XX Synthetic peptides p1-p15 (see W83043-57) are partial sequences of
 CC amino acids 1-157 of the extracellular domain of human Fas, with
 CC between 9 and 11 amino acid residues overlapping one another. p16
 CC (see W83058) is a negative control having no homology with human
 CC Fas. p1-p16 were used in an ELISA, which demonstrated that novel
 CC murine anti-human Fas monoclonal antibody HFE7A specifically binds
 CC an amino acid sequence contained in p11. The epitope (see W83030)
 CC was subsequently identified. The invention provides humanised
 CC HFE7A antibodies (see W83031-37) produced by CDR grafting. These
 CC antibodies are capable of inducing apoptosis in abnormal cells
 CC expressing Fas, and of inhibiting Fas-induced apoptosis in normal
 CC cells. They are used to evaluate, in animal models, treatments of
 CC diseases that involve Fas/Fas ligand interactions, and also to treat
 CC such diseases.
 CC
 XX
 XX Sequence 20 AA;
 SQ
 Query Match 44.7%; Score 42; DB 19; Length 20;
 Best Local Similarity 66.7%; Pred. NO. 3.5;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 3 KCKPNFYTN 11
 : : : : : :
 Db 3 rckpnfctn 11
 RESULT 9
 ID W90912 standard; peptide; 20 AA.
 XX W90912:
 AC
 XX W90912:
 AC
 XX 08-AUG-2000 (first entry)
 DT
 XX Human Fas extracellular domain antigenic peptide #12.
 DE

XX Fas; antibody; human; anti-inflammatory; anti-anemic; antidiabetic;
 KW anti-allergic; anti-arthritis; antiviral; immunomodulatory; cardiac;
 KW dermatological; immunosuppressive; thyromimetic; antirheumatic; anti-Fas;
 KW nephrotropic; antifertility; neuroprotective; antirheosclerotic;
 KW hepatotropic; humanized; apoptosis; systemic lupus erythematosus;
 KW Hashimoto disease; rheumatoid arthritis; graft versus host disease;
 KW Sjogren's syndrome; anemia; Addison's disease; scleroderma; sterility;
 KW Goodpasture syndrome; Crohn's disease; thrombopenia purpura; allergy;
 KW multiple sclerosis; Basedow's disease; myasthenia gravis;
 KW insulin dependent diabetes mellitus; arteriosclerosis; myocarditis;
 KW cardiomyopathy; glomerulonephritis; hepatitis; transplant rejection.
 XX
 OS Homo sapiens.
 XX
 PN EP990663-A2.
 XX
 PD 05-APR-2000.
 XX
 PF 29-SEP-1999; 99EP-0307711.
 XX
 PR 30-SEP-1998; 98JP-0276881.
 PR 30-SEP-1998; 98JP-0276882.
 XX
 PA (SANY) SANKYO CO LTD.
 PI Serizawa N, Haruyama H, Nakahara K, Tamaki I, Takahashi T;
 XX WPI; 2000-258930/23.
 XX
 PT New humanized anti-Fas antibody, useful for treating or preventing e.g.
 PT inflammatory or autoimmune disease, induces apoptosis selectively in
 PT cells with abnormal Fas-Fas ligand systems
 XX
 PS Example reference 6; Page 110; 263pp; English.
 CC This invention describes a novel humanized anti-Fas antibody-like
 CC molecule (I) that, induces apoptosis in cells with an abnormal Fas/Fas
 CC ligand system, by binding to Fas on the cell surface, and prevents
 CC apoptosis in cells with a normal system, by inhibiting binding between
 CC Fas and its ligand. The products of the invention have anti-inflammatory,
 CC anti-anemic, antidiabetic, anti-allergic, anti-arthritis, antiviral,
 CC immunomodulatory, dermatological, immunosuppressive, thyromimetic,
 CC antirheumatic, nephrotropic, antifertility, neuroprotective,
 CC antirheosclerotic, cardiac and hepatropic activity. (I) induce
 CC apoptosis by binding to cell surface Fas or inhibit it by competitive
 CC inhibition of ligand binding. (I) are used to treat and/or prevent
 CC diseases associated with the Fas/Fas ligand system, especially systemic
 CC lupus erythematosus, Hashimoto disease, rheumatoid arthritis, graft
 CC versus host disease, Sjogren's syndrome, pernicious or hypoplastic
 CC anemia, Addison's disease, scleroderma, Goodpasture syndrome, Crohn's
 CC disease, autoimmune hemolytic anemia, sterility, myasthenia gravis,
 CC multiple sclerosis, Basedow's disease, thrombopenia purpura, insulin
 CC dependent diabetes mellitus, allergy, arteriosclerosis, myocarditis,
 CC cardiomyopathy, glomerulonephritis, hepatitis (fulminant, chronic, viral
 CC (B, C or D) or alcoholic), and transplant rejection. (I) selectively
 CC inhibit apoptosis in normal cells but selectively induce it in abnormal
 CC cells. They bind to both human and murine Fas, so can be evaluated in
 CC murine disease models. (I) act on the active site of Fas, i.e. they mimic
 CC the native ligand, do not induce liver disease, and have reduced risk of
 CC inducing a human anti-murine antibody response. This sequence represents
 CC a human Fas extracellular domain derived antigenic peptide described in
 CC the method of the invention.
 XX
 SO Sequence 20 AA;

Query Match 44.7%; Score 42; DB 21; Length 20;
 Best Local Similarity 66.7%; Pred. No. 3.5;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 KCKPNEYNN 11
 :|||||: |

Db 3 rckpnffcn 11
 RESULT 10
 Y69518
 ID Y69518 standard; peptide; 20 AA.
 XX
 AC Y69518;
 XX
 PT 10-APR-2000 (first entry)
 XX
 DE Human Fas peptide fragment, Fp12.
 XX
 KW Fp12; Fas fragment; APO-1; CD95; anti-Fas autoantibody; antibody;
 KW modulator; apoptosis; proliferation.
 XX
 OS Homo sapiens.
 XX
 PN WO9965935-A2.
 XX
 PD 23-DEC-1999.
 XX
 PF 15-JUN-1999; 99WO-EP04105.
 XX
 PR 18-JUN-1998; 98GB-0013194.
 PR 12-MAR-1999; 99GB-0005793.
 XX
 PA (KARO-) KAROLINSKA INNOVATIONS AB.
 PI Chlodt F;
 XX
 PT WPI; 2000-106082/09.
 XX
 PT Fas peptide fragments useful for the treatment of proliferative
 PT disorders
 XX
 PS Claim 3; Page 57; 71pp; English.
 CC Sequences Y69514-Y69521 represent peptide fragments of the Fas protein
 CC (also known as APO-1 or CD95). Fas is a type I cellular receptor which
 CC transduces an apoptotic signal on binding of its ligand, Fas ligand
 CC (FasL). Fas autoantibodies have been isolated from the serum of healthy
 CC blood donors - these may represent an additional mode of regulation of
 CC Fas-mediated signals in vivo. The Fas peptides may be used to modulate
 CC apoptosis via blocking autoantibody binding to Fas, thereby reducing or
 CC increasing FasL binding, which results in inhibition or stimulation of
 CC apoptosis. Pp5 (Y69514) is thus able to induce apoptosis while Fp11
 CC (Y69517) and Fp17 (Y69519) are able to block apoptosis. Fp8 (Y69515) and
 CC Fp9 (Y69516) comprise amino acids which are important for binding of Fas
 CC to FasL. The Fas peptides can be used to obtain antibodies containing a
 CC Fas binding site. The peptides and antibodies can be used in assay
 CC methods to obtain candidate modulators of Fas-mediated apoptosis. The
 CC candidate modulator may also modulate cellular proliferation. The
 CC peptides, mixtures of peptides, nucleic acids or antibodies are useful
 CC for methods of treating proliferative disorders. The disorders that may
 CC be prevented or treated include tumours, cancer, psoriasis, type I
 CC diabetes, multiple sclerosis, liver cirrhosis and HIV infection.
 XX
 SO Sequence 20 AA;

Query Match 44.7%; Score 42; DB 21; Length 20;
 Best Local Similarity 66.7%; Pred. No. 3.5;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 KCKPNEYNN 11
 :|||||: |

RESULT 11
 R27396
 ID R27396 standard; peptide; 58 AA.

XX	R27396;
AC	20-MAY-1998 (first entry)
DT	Human neutrophil elastase inhibitor BITI-E7-1222.
DE	Kunitz Domain serine protease inhibitor; mutant:
XX	Bovine pancreatic trypsin inhibitor.
KW	Synthetic.
OS	
XX	
FH	Key
FT	Location/Qualifiers
FT	region 1..4
FT	/note= "mutated from ITI-D1 (KEDS) -> (RPDF)"
FT	region 11
FT	/note= "mutated from ITI-D1 (A) -> (T)"
FT	region 15..19
FT	/note= "mutated from ITI-D1 (MOMTS) -> (VAMPP)"
PN	W09215605-A.
XX	
PD	17-SEP-1992.
XX	
PF	28-FEB-1992; 92MO-US01501.
XX	
PR	01-MAR-1991; 91US-0664989.
PR	17-JUN-1991; 91US-0715834.
XX	
PA	(PROT-) PROTEIN ENG CORP.
XX	Guterman SK, Kent RB, Ladner RC, Ley AC, Markland W, Roberts BL;
PI	WI; 1992-331666/40.
DR	
XX	
PT	New peptide inhibitors of elastase or cathepsin G - are e.g.
PT	mutants of Kunitz Domain serine protease inhibitors, useful for
PT	treating and preventing conditions caused by excessive neutrophil
PT	elastase or cathepsin G
PS	Example; Page 103; 126pp; English.
XX	
CC	The sequence is that of a mutant of inter-alpha trypsin inhibitor-D1
CC	(ITI-D1) BITI-E7-1222 which is an inhibitor with strong (Kd=10-9 to
CC	10-11) specific binding activity for human neutrophil elastase (hNE).
CC	It can be used for the treatment or prophylaxis of a condition caused by
CC	excessive hNE activity, e.g. inflammation, emphysema, cystic fibrosis,
CC	adult respiratory distress syndrome or rheumatoid arthritis. It may
CC	also be used to purify hNE. See also R27373-R27403 and R27443.
SO	Sequence 58 AA;
OY	Query Match 44.7%; Score 42; DB 13; Length 58;
DB	Best Local Similarity 50.0%; Pred. No. 10;
	Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
	1 CVKCKPNFYNGGS 14
	::
	14 cvamfryfyngts 27
RESULT 12	
ID R27397	
XX R27397 standard; peptide; 58 AA.	
AC R27397;	
XX	
DT 20-MAY-1998 (first entry)	
XX	
DE Human neutrophil elastase inhibitor AMIMOL.	
XX	
KW Kunitz Domain serine protease inhibitor; mutant;	
PW	

KV	Bovine pancreatic trypsin inhibitor.
XX	Synthetic.
OS	
FH	Key
FT	region
FT	/note= "mutated from ITI-DL (S) -> (F)"
FT	region
FT	15..19
FT	/note= "mutated from ITI-DL (MGMTS) -> (VAMFP)"
FT	
PN	W09215605-A.
XX	
PD	17-SEP-1992.
XX	
PF	28-FEB-1992; 92WO-US01501.
XX	
PR	01-MAR-1991; 91US-0664989.
PR	17-JUN-1991; 91US-0715834.
XX	
PA	(PROT-) PROTEIN ENG CORP.
XX	
PI	Guterman SK, Kent RB, Ladner RC, Ley AC, Markland W, Roberts BL;
XX	
DR	WPI: 1992-331666/40.
XX	
PT	New peptide inhibitors of elastase or cathepsin G - are e.g.
PT	mutants of Kunitz Domain serine protease inhibitors, useful for
PT	treating and preventing conditions caused by excessive neutrophil
PT	elastase or cathepsin G
XX	
PS	Example: Page 103; 126pp; English.
XX	
CC	The sequence is that of a mutant of inter-alpha trypsin inhibitor-D1
CC	(ITI-D1) AMINO1 which is an inhibitor with strong (Kd=10 ⁻⁹ to 10 ⁻¹¹)
CC	specific binding activity for human neutrophil elastase (hNE). It can
CC	be used for the treatment or prophylaxis of a condition caused by
CC	excessive hNE activity, e.g. inflammation, emphysema, cystic fibrosis,
CC	adult respiratory distress syndrome or rheumatoid arthritis. It may
CC	also be used to purify hNE. See also R27373-R27403 and R27443.
XX	
SO	Sequence 58 AA:
OY	Query Match 44.7%; Score 42; DB 13; Length 58; Best Local Similarity 50.0%; Pred. No. 10; Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0. 1 CVCKENFYNGGS 14 : DQ 14 cvamfpyrlyngts 27
RESULT 13	
R27400	
ID	R27400 standard; peptide: 58 AA.
XX	R27400;
DJ	20-MAY-1998 (first entry)
XX	
DE	Human neutrophil elastase inhibitor BII-E7-141.
XX	
KW	Kunitz Domain serine protease inhibitor; mutant;
KW	Bovine pancreatic trypsin inhibitor.
XX	
OS	Synthetic.
XX	
FH	Key
FT	region
FT	/note= "mutated from ITI-DL (KEDS) -> (RPDF)"
FT	region
FT	15..19
FT	/note= "mutated from ITI-DL (MGMTS) -> (VAMFP)"
FT	31..34

FT /note= "mutated from ITI-D1 (ETFO) -> (QTEV)"
XX
PN W09215605-A.
XX
PD 17-SEP-1992.
XX
PE 28-FEB-1992; 92WO-US01501.
XX
PR 01-MAR-1991; 91US-0664989.
PR 17-JUN-1991; 91US-0715834.
XX
PA (PROT-) PROTEIN ENG CORP.
XX
PI Guterman SK, Kent RB, Ladner RC, Ley AC, Markland W, Roberts BL.
XX
DR WPI; 1992-331666/40.
XX
PT New peptide inhibitors of elastase or cathepsin G - are e.g.
PT mutants of Kunitz Domain serine protease inhibitors, useful for
PT treating and preventing conditions caused by excessive neutrophil
PT elastase or cathepsin G
XX
PS Example; Page 103; 126pp; English.
XX
CC The sequence is that of a mutant of inter-alpha trypsin inhibitor-D1
CC (ITI-D1) BIRI-E7-141 which is an inhibitor with very strong (Kd< 10-11)
CC specific binding activity for human neutrophil elastase (hNE). It can
CC be used for the treatment or prophylaxis of a condition caused by
CC excessive hNE activity, e.g. inflammation, emphysema, cystic fibrosis,
CC adult respiratory distress syndrome or rheumatoid arthritis. It may
CC also be used to purify hNE. See also R27373-R27403 and R27443.
XX
SQ Sequence 58 AA:

QY 1 CVCKCKPNFYNGGS 14
 || |::||| |
 14 cvamfprrfyngts 27
Db

RESULT 14
R27402
ID R27402 standard; peptide; 58 AA.
XX
AC R27402;
XX
DT 20-MAY-1998 (first entry)
XX
DE Human neutrophil elastase inhibitor MUTOE.
XX
KW Kunitz Domain serine protease inhibitor; mutant;
KW Bovine pancreatic trypsin inhibitor.
XX
OS Synthetic.
OS
XX
FH Key
FH region 1..4 Location/Qualifiers
FT /note= "mutated from ITI-D1 (KEDS) -> (RPDF)"
FT region 15..19
FT /note= "mutated from ITI-D1 (MGMS) -> (VAMFP)"
FT region 34
FT /note= "mutated from ITI-D1 (Q) -> (V)"
XX
XX W09215605-A.
XX
PD 17-SEP-1992.
XX
PE 28-FEB-1992; 92WO-US01501.
XX
PT

PR 01-MAR-1991; 91US-0664989.
PR 17-JUN-1991; 91US-0715834.
XX
PA (PROT-) PROTEIN ENG CORP.
XX
PI Guterman SK, Kent RB, Ladner RC, Ley AC, Markland W, Roberts BL.
XX
DR WPI; 1992-331666/40.
XX
PT New peptide inhibitors of elastase or cathepsin G - are e.g.
PT mutants of Kunitz Domain serine protease inhibitors, useful for
PT treating and preventing conditions caused by excessive neutrophil
PT elastase or cathepsin G
XX
PS Example; Page 103; 126pp; English.
XX
CC The sequence is that of a mutant of inter-alpha trypsin inhibitor-D1
CC (ITI-D1) MUTOE which is an inhibitor with very strong (Kd< 10-11)
CC specific binding activity for human neutrophil elastase (hNE). It can
CC be used for the treatment or prophylaxis of a condition caused by
CC excessive hNE activity, e.g. inflammation, emphysema, cystic fibrosis,
CC adult respiratory distress syndrome or rheumatoid arthritis. It may
CC also be used to purify hNE. See also R27373-R27403 and R27443.
XX
SQ Sequence 58 AA:

QY 1 CVCKCKPNFYNGGS 14
 || |::||| |
 14 cvamfprrfyngts 27
Db

RESULT 15
R27394
ID R27394 standard; peptide; 58 AA.
XX
AC R27394;
XX
DT 20-MAY-1998 (first entry)
XX
DE Human neutrophil elastase inhibitor ITI-E7.
XX
KW Kunitz Domain serine protease inhibitor; mutant;
KW Bovine pancreatic trypsin inhibitor.
XX
OS Synthetic.
OS
XX
FH Key
FH region 15..19 Location/Qualifiers
FT /note= "mutated from ITI-D1 (MGMS) -> (VAMFP)"
FT region 34
FT /note= "mutated from ITI-D1 (Q) -> (V)"
XX
XX W09215605-A.
XX
PD 17-SEP-1992.
XX
PE 28-FEB-1992; 92WO-US01501.
XX
PR 01-MAR-1991; 91US-0664989.
PR 17-JUN-1991; 91US-0715834.
XX
PA (PROT-) PROTEIN ENG CORP.
XX
PI Guterman SK, Kent RB, Ladner RC, Ley AC, Markland W, Roberts BL.
XX
DR WPI; 1992-331666/40.
XX
PT New peptide inhibitors of elastase or cathepsin G - are e.g.
PT mutants of Kunitz Domain serine protease inhibitors, useful for
PT treating and preventing conditions caused by excessive neutrophil

PT elastase or cathepsin G
 XX
 PS Example; Page 103; 126pp; English.
 XX
 CC The sequence is that of a mutant of inter-alpha trypsin inhibitor-D1
 CC (ITI-D1) ITI-E7 which is an inhibitor with moderate (Kd 10-8 to 10-9)
 CC specific binding activity for human neutrophil elastase (hNE). It can
 CC be used for the treatment or prophylaxis of a condition caused by
 CC excessive hNE activity, e.g. inflammation, emphysema, cystic fibrosis,
 CC adult respiratory distress syndrome or rheumatoid arthritis. It may
 CC also be used to purify hNE. See also R27373-R27403 and R27443.
 XX
 SO Sequence 58 AA;

Query Match 44.7%; Score 42; DB 13; Length 58;
 Best Local Similarity 50.0%; Pred. No. 10;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 1 CVCKPNEFYNGS 14
 || |::||| |
 Db 14 cvamfpyfyngts 27

Search completed: March 6, 2001, 12:49:35
 Job time: 96 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:53:49 ; Search time: 57.76 Seconds
(without alignments)
4.663 Million cell updates/sec

Title: US-09-196-161d-6
Perfect score: 94
Sequence: 1 CVKCKRNFYNGSP 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues
Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA: *
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6.COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	46.8	58	1	US-08-358-160-18
2	43	45.7	23	5	5196510-19
3	43	45.7	335	3	US-08-815-469-6
4	42	44.7	58	1	US-08-358-160-9
5	42	44.7	58	1	US-08-358-160-10
6	42	44.7	58	1	US-08-358-160-11
7	42	44.7	58	1	US-08-358-160-12
8	42	44.7	58	1	US-08-358-160-15
9	42	44.7	58	1	US-08-358-160-16
10	42	44.7	58	1	US-08-358-160-17
11	42	44.7	119	2	US-08-219-237B-3
12	42	44.7	219	3	US-08-974-022-45
13	42	44.7	243	3	US-08-460-309-15
14	42	44.7	243	2	US-08-460-309-16
15	42	44.7	243	2	US-08-125-077-15
16	42	44.7	243	2	US-08-125-077-16
17	42	44.7	314	1	US-08-444-231-19
18	42	44.7	314	1	US-08-152-443A-19
19	42	44.7	314	2	US-08-460-309-19
20	42	44.7	314	2	US-08-125-077-19
21	42	44.7	314	4	PCT-US95-17083-4
22	42	44.7	335	2	US-08-219-237B-2
23	42	44.7	335	2	US-08-409-338-1
24	42	44.7	335	4	PCT-US95-17083-2
25	42	44.7	2556	1	US-08-083-590A-20
26	42	44.7	2556	3	US-08-532-384-20
27	42	44.7	3075	2	US-08-460-309-5
28	42	44.7	3075	2	US-08-125-077-5

29	41	43.6	65	1	US-08-358-160-92	Sequence 92, Appl
30	41	43.6	65	5	546783-12	Patent No. 546783
31	41	43.6	197	2	US-08-505-606-1	Sequence 1, Appl
32	41	43.6	225	1	US-08-152-019A-33	Sequence 33, Appl
33	41	43.6	1286	5	5206163-1	Patent No. 5206163
34	41	43.6	2523	1	US-08-185-432-18	Sequence 18, Appl
35	40	42.6	57	1	US-08-358-160-103	Sequence 103, App
36	40	42.6	57	1	US-08-358-160-108	Sequence 108, App
37	40	42.6	1251	4	PCT-US95-02251-3	Sequence 3, Appl
38	40	42.6	1252	1	US-08-199-780-3	Sequence 3, Appl
39	40	42.6	1252	2	US-08-316-650-3	Sequence 3, Appl
40	40	42.6	1253	3	US-08-479-722B-4	Sequence 4, Appl
41	39.5	42.0	33	5	5196510-4	Patent No. 5196510
42	39.5	42.0	3111	2	US-08-460-309-4	Sequence 4, Appl
43	39.5	42.0	3111	2	US-08-125-077-4	Sequence 4, Appl
44	39	41.5	207	3	US-08-974-022-47	Sequence 47, Appl
45	39	41.5	219	1	US-08-152-019A-31	Sequence 31, Appl

ALIGNMENTS

RESULT 1
US-08-358-160-18
Sequence 18, Application US/08358160
Patent No. 5663143
GENERAL INFORMATION:
APPLICANT: LEY, Arthur C.
APPLICANT: LADNER, Robert C.
APPLICANT: GUTERMAN, Sonia K.
APPLICANT: ROBERTS, Bruce L.
APPLICANT: MARKLAND, William
APPLICANT: KENT, Rachel B.
TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
TITLE OF INVENTION: DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
NUMBER OF SEQUENCES: 234
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W. Suite 300
CITY: Washington
STATE: District of Columbia
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/358,160
FILING DATE: 16-DEC-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/133,031
FILING DATE: 13-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/009,319
FILING DATE: 26-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/664,989
FILING DATE: 01-MAR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/487,063
FILING DATE: 02-MAR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/240,160
FILING DATE: 02-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Cooper, Iver P.
REGISTRATION NUMBER: 28,005
REFERENCE/DOCKET NUMBER: LEY-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197

```

? TELEFAX: 202-737-3528
?
? TELEX: 248633
? INFORMATION FOR SEQ ID NO:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 58 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
?
? US-08-358-160-18

```

Query Match	46.88;	Score 44;	DB 1;	Length 58
Best local Similarity	50.0%;	Pred. No. 3.5;		
Matches	7;	Conservative	2;	Mismatches 5; Indels

RESULT 2
5196510-19
Patent No. 5196510
APPLICANT: ROOMWELL, JOHN D.; MCKEARN, THOMAS J.; ALVAREZ,
VERNON, L.; RADCLIFFE, ROBERT D.
TITLE OF INVENTION: MOLECULAR RECOGNITION UNITS
NUMBER OF SEQUENCES: 43
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/519,702
FILING DATE: 07-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 291,730
FILING DATE: 29-DEC-1988
SEQ ID NO: 19:
LENGTH: 23
5196510-19

Query Match	45.7%	Score 43	DB 5	length 23
Best Local Similarity	40.0%	Pred. NO. 1.9		
Matches 6	Conservative 3	Mismatches 6	Indels 0	Gaps 0

RESULT 3
US-08-815-469-6
Sequence 6, Application US/08815469
Patent No. 6153402
GENERAL INFORMATION:
APPLICANT: Yu, Guo-Liang
APPLICANT: Ni, Jian
APPLICANT: Dixit, Vishva
APPLICANT: Gentz, Reiner L.
TITLE OF INVENTION: Death Domain Containing Receptors
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSER: Sterne, Kessler, Goldstein & Fox, P.L.L.C
STREET: 1100 New York Ave., NW, Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

```

1  APPLICATION NUMBER: US/08/815,469
2  FILING DATE: HERMITH
3  CLASSIFICATION: 435
4  PRIOR APPLICATION DATA:
5  APPLICATION NUMBER: No. 6153402 Yet Assigned
6  FILING DATE: 06-FEB-1997
7  PRIOR APPLICATION DATA:
8  APPLICATION NUMBER: US 60/028,711
9  FILING DATE: 17-OCT-1996
10 PRIOR APPLICATION DATA:
11 APPLICATION NUMBER: US 60/013,285
12 FILING DATE: 12-MAR-1996
13 ATTORNEY/AGENT INFORMATION:
14 NAME: Seife, Eric K.
15 REGISTRATION NUMBER: 36,688
16 REFERENCE/DOCKET NUMBER: 1488. 0310003/EKS/KRM
17 TELECOMMUNICATION INFORMATION:
18 TELEPHONE: 202-371-2600
19 TELEFAX: 202-371-2540
20 INFORMATION FOR SEQ ID NO: 6:
21 SEQUENCE CHARACTERISTICS:
22 LENGTH: 335 amino acids
23 TYPE: amino acid
24 STRANDEDNESS: not relevant
25 TOPOLOGY: not relevant
26 MOLECULE TYPE: protein
27 US-08-815-469-6

```

Query Match	45.7%	Score 43	DB 3	length 335
Best Local Similarity	66.7%	Pred. NO. 29		
Matches	6	Conservative	2	Mismatches 1; Indels 0; Gaps 0

RESULT 4
 US-08-358-160-9
 Sequence 9, Application US/08358160
 Patent No. 5663143
 GENERAL INFORMATION:
 APPLICANT: LEY, Arthur C.
 APPLICANT: LADNER, Robert C.
 APPLICANT: GUTERMAN, Sonia K.
 APPLICANT: ROBERTS, Bruce L.
 APPLICANT: MARKLAND, William
 APPLICANT: KENT, Rachel B.
 TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
 TITLE OF INVENTION: DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
 NUMBER OF SEQUENCES: 234
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BROWDY AND NEIMARK
 STREET: 419 Seventh Street, N.W. Suite 300
 City: Washington
 STATE: District of Columbia
 COUNTRY: USA
 Zip: 20004
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/358,160
 FILING DATE: 16-DEC-1994
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/133,031
 FILING DATE: 13-OCT-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/009,319

FILED DATE: 26-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/664,989
FILING DATE: 01-MAR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/487,063
FILING DATE: 02-MAR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/240,160
FILING DATE: 02-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Cooper, Iver P.
REGISTRATION NUMBER: 28,005
REFERENCE/DOCKET NUMBER: LEY-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 58 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-358-160-9

Query Match 44.7% Score 42: DB 1: Length 58:
Best Local Similarity 50.0% Pred. No. 6.9;
Matches 7: Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 CVCKRPNFYNGS 14
|||::|||
DB 14 CVAMPFRFYNGTS 27

RESULT 5
US-08-358-160-10
Sequence 10, Application US/08358160
Patent No. 5663143
GENERAL INFORMATION:
APPLICANT: LEY, Arthur C.
APPLICANT: LADNER, Robert C.
APPLICANT: GUTERMAN, Sonia K.
APPLICANT: ROBERTS, Bruce L.
APPLICANT: MARKLAND, William
APPLICANT: KENT, Rachel B.
TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
TITLE OF INVENTION: DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
NUMBER OF SEQUENCES: 234
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W. Suite 300
CITY: Washington
STATE: District of Columbia
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/358,160
FILING DATE: 16-DEC-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/133,031
FILING DATE: 13-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/009,319
FILING DATE: 26-JAN-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/664,989
FILING DATE: 01-MAR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/487,063
FILING DATE: 02-MAR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/240,160
FILING DATE: 02-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Cooper, Iver P.
REGISTRATION NUMBER: 28,005
REFERENCE/DOCKET NUMBER: LEY-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 58 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-358-160-10

Query Match 44.7% Score 42: DB 1: Length 58:
Best Local Similarity 50.0% Pred. No. 6.9;
Matches 7: Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 CVCKRPNFYNGS 14
|||::|||
DB 14 CVAMPFRFYNGTS 27

RESULT 6
US-08-358-160-11
Sequence 11, Application US/08358160
Patent No. 5663143
GENERAL INFORMATION:
APPLICANT: LEY, Arthur C.
APPLICANT: LADNER, Robert C.
APPLICANT: GUTERMAN, Sonia K.
APPLICANT: ROBERTS, Bruce L.
APPLICANT: MARKLAND, William
APPLICANT: KENT, Rachel B.
TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
TITLE OF INVENTION: DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
NUMBER OF SEQUENCES: 234
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W. Suite 300
CITY: Washington
STATE: District of Columbia
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/358,160
FILING DATE: 16-DEC-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/133,031
FILING DATE: 13-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/009,319
FILING DATE: 26-JAN-1993
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/664,989
FILING DATE: 01-MAR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/487,063
FILING DATE: 02-MAR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/240,160
FILING DATE: 02-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Cooper, Iver P.
REGISTRATION NUMBER: 28,005
REFERENCE/DOCKET NUMBER: LEY-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 58 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-358-160-11

Query Match 44.7%; Score 42; DB 1; Length 58;
Best Local Similarity 50.0%; Pred. No. 6.9;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CVKCKPNEYNGS 14
11 | :||| |
Db 14 CVAMFPFYNGTS 27

RESULT 7
US-08-358-160-12
Sequence 12, Application US/08358160
Patent No. 5663143
GENERAL INFORMATION:
APPLICANT: LEY, Arthur C.
APPLICANT: LADNER, Robert C.
APPLICANT: GUTERMAN, Sonia K.
APPLICANT: ROBERTS, Bruce L.
APPLICANT: MARKLAND, William
APPLICANT: KENT, Rachel B.
TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
TITLE OF INVENTION: DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
NUMBER OF SEQUENCES: 234
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W. Suite 300
CITY: Washington
STATE: District of Columbia
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/358,160
FILING DATE: 16-DEC-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/133,031
FILING DATE: 13-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/009,319
FILING DATE: 26-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/664,989

FILING DATE: 01-MAR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/487,063
FILING DATE: 02-MAR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/240,160
FILING DATE: 02-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Cooper, Iver P.
REGISTRATION NUMBER: 28,005
REFERENCE/DOCKET NUMBER: LEY-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 58 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-358-160-12

Query Match 44.7%; Score 42; DB 1; Length 58;
Best Local Similarity 50.0%; Pred. No. 6.9;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CVKCKPNEYNGS 14
11 | :||| |
Db 14 CVAMFPFYNGTS 27

RESULT 8
US-08-358-160-15
Sequence 15, Application US/08358160
Patent No. 5663143
GENERAL INFORMATION:
APPLICANT: LEY, Arthur C.
APPLICANT: LADNER, Robert C.
APPLICANT: GUTERMAN, Sonia K.
APPLICANT: ROBERTS, Bruce L.
APPLICANT: MARKLAND, William
APPLICANT: KENT, Rachel B.
TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
TITLE OF INVENTION: DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
NUMBER OF SEQUENCES: 234
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W. Suite 300
CITY: Washington
STATE: District of Columbia
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/358,160
FILING DATE: 16-DEC-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/133,031
FILING DATE: 13-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/009,319
FILING DATE: 26-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/664,989
FILING DATE: 01-MAR-1991

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/487,063
FILING DATE: 02-MAR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/240,160
FILING DATE: 02-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Cooper, Iver P.
REGISTRATION NUMBER: 28,005
REFERENCE/DOCKET NUMBER: LEY=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 58 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-358-160-15

Query Match 44.7%; Score 42; DB 1; Length 58;
Best Local Similarity 50.0%; Pred. No. 6.9;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 CVKCKPNFYNGGS 14
DB 14 CVAMPFRYFNGTS 27

RESULT 9
US-08-358-160-16
Sequence 16, Application US/08358160
Patent No. 5663143
GENERAL INFORMATION:
APPLICANT: LEY, Arthur C.
APPLICANT: LADNER, Robert C.
APPLICANT: GUTERMAN, Sonia K.
APPLICANT: ROBERTS, Bruce L.
APPLICANT: MARKLAND, William
APPLICANT: KENT, Rachel B.
TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
TITLE OF INVENTION: DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
NUMBER OF SEQUENCES: 234
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W. Suite 300
CITY: Washington
STATE: District of Columbia
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/358,160
FILING DATE: 16-DEC-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/133,031
FILING DATE: 13-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/009,319
FILING DATE: 26-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/664,989
FILING DATE: 01-MAR-1991
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/487,063
FILING DATE: 02-MAR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/240,160
FILING DATE: 02-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Cooper, Iver P.
REGISTRATION NUMBER: 28,005
REFERENCE/DOCKET NUMBER: LEY=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 58 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-358-160-16

Query Match 44.7%; Score 42; DB 1; Length 58;
Best Local Similarity 50.0%; Pred. No. 6.9;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 CVKCKPNFYNGGS 14
DB 14 CVAMPFRYFNGTS 27

RESULT 10
US-08-358-160-17
Sequence 17, Application US/08358160
Patent No. 5663143
GENERAL INFORMATION:
APPLICANT: LEY, Arthur C.
APPLICANT: LADNER, Robert C.
APPLICANT: GUTERMAN, Sonia K.
APPLICANT: ROBERTS, Bruce L.
APPLICANT: MARKLAND, William
APPLICANT: KENT, Rachel B.
TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
TITLE OF INVENTION: DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
NUMBER OF SEQUENCES: 234
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W. Suite 300
CITY: Washington
STATE: District of Columbia
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/358,160
FILING DATE: 16-DEC-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/133,031
FILING DATE: 13-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/009,319
FILING DATE: 26-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/664,989
FILING DATE: 01-MAR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/487,063

ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460.309
FILING DATE: 22-SEP-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/125.077
FILING DATE: 22-SEP-1993
APPLICATION NUMBER: US PCT/US 94/10730
FILING DATE: 21-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/472.319
FILING DATE: 30-JAN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/919.951
FILING DATE: 27-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 9721
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 243 amino acids
TYPE: amino acid
TOPOLOGY: linear

US-08-460-309-15

Query Match 44.7%; Score 42; DB 2; Length 243;
Best Local Similarity 66.7%; Pred. No. 29;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVKCKPNEY 9
Db 213 CDRCKPGFY 221

RESULT 14
US-08-460-309-16
Sequence 16, Application US/08460309
Patent No. 5837496
GENERAL INFORMATION:
APPLICANT: Engvall, Eva
APPLICANT: Leiyo, Ilmo
TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/460.309
FILING DATE: 22-SEP-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/125.077
FILING DATE: 22-SEP-1993
APPLICATION NUMBER: US PCT/US 94/10730
FILING DATE: 21-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/472.319
FILING DATE: 30-JAN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/919.951
FILING DATE: 27-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 9721
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 243 amino acids
TYPE: amino acid
TOPOLOGY: linear

US-08-460-309-16

Query Match 44.7%; Score 42; DB 2; Length 243;
Best Local Similarity 66.7%; Pred. No. 29;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVKCKPNEY 9
Db 213 CDRCKPGFY 221

RESULT 15
US-08-125-077-15
Sequence 15, Application US/08125077
Patent No. 5872231
Patent No. 5872231 5840863
GENERAL INFORMATION:
APPLICANT: Engvall, Eva
APPLICANT: Leiyo, Ilmo
TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/125.077
FILING DATE: 22-SEP-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US 94/10730
FILING DATE: 21-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/472.319
FILING DATE: 30-JAN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/919.951

FILING DATE: 27-JUL-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Campbell, Cathryn A.
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-LA 9721
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 15:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 243 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-125-077-15

Query Match 44.7%; Score 42; DB 2; Length 243;
 Best Local Similarity 66.7%; Pred. No. 29;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 CVCKRPNEY 9
 I : I I I I I
 Db 213 CDRCKPGFY 221

Search completed: March 6, 2001, 12:53:50
 Job time: 350 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:50:52 ; Search time 70.34 Seconds
(without alignments)
14.480 Million cell updates/sec

Title: US-09-196-161d-6
Perfect score: 94
Sequence: 1 CVKCKPNEYNGSP 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_66:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	94	100.0	395	2 A46031	immobilization sur
2	56	59.6	677	2 C42125	trophozoite cystei
3	48	51.1	3672	2 T23433	hypothetical prote
4	48	51.1	3704	2 T37316	probable laminin a
5	47	50.0	392	2 S37902	hypothetical prote
6	45	47.9	335	2 T20465	hypothetical prote
7	45	47.9	1571	2 T13711	polypeptide - frui
8	44	46.8	305	2 T12744	pseudomurein endo
9	44	46.8	342	2 T30370	hypothetical prote
10	44	46.8	2102	2 T15626	hypothetical prote
11	43	45.7	139	2 T15626	extensin precursor
12	43	45.7	201	2 S74667	hypothetical prote
13	43	45.7	287	2 H69414	conserved hypotet
14	43	45.7	400	2 T46383	hypothetical prote
15	43	45.7	1607	1 MMMSB2	laminin gamma-1 ch
16	43	45.7	1609	1 MMMSB2	laminin gamma-1 ch
17	42.5	45.2	393	2 B72455	threonine synthase
18	42	44.7	242	2 A71967	hypothetical prote
19	42	44.7	314	2 I37383	F45 soluble protei
20	42	44.7	335	2 A40036	apoptosis-mediatin
21	42	44.7	390	2 E75586	urea/hot-chain a
22	42	44.7	652	2 S71753	repellent protein
23	42	44.7	785	2 T00474	hypothetical prote
24	42	44.7	803	2 S45916	hypothetical prote
25	42	44.7	2531	2 S18188	notch protein homo
26	42	44.7	2531	2 A46019	Notch-1 protein -
27	42	44.7	2555	2 A40043	notch protein homo
28	42	44.7	2946	2 T15840	hypothetical prote
29	42	44.7	3075	2 S14458	laminin alpha-1 ch

30	42	44.7	3084	1 MMMSA	laminin alpha-1 ch
31	42	44.7	3712	2 S18253	laminin alpha-1 ch
32	41.5	44.1	1291	2 T21694	hypothetical prote
33	41	43.6	65	1 T1YIVC	venom basic protei
34	41	43.6	367	2 T22569	hypothetical prote
35	41	43.6	435	2 I54182	tumor necrosis fac
36	41	43.6	455	2 S71344	purinergic recepto
37	41	43.6	596	2 A45664	variant-specific s
38	41	43.6	713	2 I65253	disintegrin-like t
39	41	43.6	1139	2 S28277	hypothetical prote
40	41	43.6	1323	2 T27682	hypothetical prote
41	41	43.6	1374	2 S70712	protein-tyrosine k
42	41	43.6	1639	1 MMFFB2	laminin gamma-1 ch
43	41	43.6	2524	2 A35844	Xotch protein - Af
44	41	43.6	3635	2 T10053	laminin alpha 5 ch
45	41	43.6	3707	2 S18252	heparan sulfate pr

ALIGNMENTS

RESULT 1
A46031
immobilization surface I-antigen precursor - Ichthyophthirius multifiliis (fragment)
C:Species: Ichthyophthirius multifiliis
C:Date: 29-Sep-1999 #sequence_revision 29-Sep-1999 #text_change 07-Dec-1999
C:Accession: A46031
R:Clark, T.G.; McGraw, R.A.; Dickerson, H.W.
Proc. Natl. Acad. Sci. U.S.A. 89, 6363-6367, 1992
A:Title: Developmental expression of surface antigen genes in the parasitic ciliate I
A:Reference number: A46031; MUID:92335298
A:Accession: A46031
A:Molecule type: mRNA; protein
A:Residues: 1-395 <CLIA>
A:Cross-references: GB:M2907; NID:93628568; PID:ANC36158.1; PID:93628569
A>Note: The authors translated the codon UUG for residue 330 as Ile
A>Note: sequence extracted from NCBI Backbone (NCBIN:108734; NCBIPI:108735); the sequ
C:Genetics:
A:Genetic code: SGC5
C:Keywords: glycoprotein; surface antigen
F:2-395/Product: immobilization surface I-antigen #status experimental <MAT>
F:156,191,245,281/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 94; DB 2; Length 395;
Best local similarity 100.0%; Pred. No. 1e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CVKCKPNEYNGSP 15
DB 108 CVKCKPNEYNGSP 122

RESULT 2
C42125
trophozoite cysteine-rich surface antigen 72 - Giardia lamblia (fragment)
N:Alternate names: CRP72
C:Species: Giardia lamblia
C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 05-Dec-1998
C:Accession: C42125
R:Adam, R.D.; Yang, Y.M.; Nash, T.E.
Mol. Cell. Biol. 12, 1194-1201, 1992
A:Title: The cysteine-rich protein gene family of Giardia lamblia: loss of the CRP170
A:Reference number: A42125; MUID:92186850
A:Accession: C42125
A:Molecule type: DNA
A:Residues: 1-677 <ADA>
A:Cross-references: GB:M83934; NID:9159123
A:Experimental source: trophozoites
A>Note: sequence extracted from NCBI backbone (NCBIN:88443; NCBIPI:88444); this ORF is
C:Keywords: surface antigen

Query Match 59.6%; Score 56; DB 2; Length 677;
 Best Local Similarity 53.3%; Pred. No. 0.76;
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 CVKCKPNFYNGSP 15
 | | | | | | | | | |
 Db 360 CTCKCPGFEMKGNCP 374

RESULT 3

T23433
 hypothetical protein K08C7.3 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000

C:Accession: T23433

R:Berks, M.
 submitted to the EMBL Data Library, March 1996

A:Reference number: Z19740

A:Accession: T23433

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3672 <NTL>

A:Cross-references: EMBL:Z70286; PTDN:CA94293.1; GSPDB:GN00022; CESP:K08C7.3

A:Experimental source: clone K08C7

C:Genetics:

A:Gene: CESP:K08C7.3

A:Map position: 4

A:Introns: 66/1; 284/3; 563/1; 1187/3; 1248/3; 1300/1; 1460/1; 1623/3; 2361/3; 2988/3; 3

C:Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like

Query Match 51.1%; Score 48; DB 2; Length 3672;
 Best Local Similarity 77.8%; Pred. No. 49;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 CVKCKPNFY 9
 | : | | | | | | | |
 Db 685 CDCKCPNPFY 693

RESULT 4

T37316
 Probable laminin alpha chain - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Jan-2000

C:Accession: T37316

R:Job, K.; Zhu, K.; Hedgecock, E.M.; Inoue, T.; Horl, K.
 submitted to the EMBL Data Library, August 1998

A:Description: laminin alpha chain gene in the nematode *C. elegans*.

A:Reference number: Z21681

A:Accession: T37316

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3704 <JOH>

A:Cross-references: EMBL:AB016806; PTDN:BA32347.1

A:Experimental source: strain N2

C:Genetics:

A:Gene: epl-1

A:Map position: IV

A:Introns: 66/1; 284/3; 563/1; 1187/3; 1248/3; 1300/1; 1460/1; 1623/3; 2361/3; 2988/3; 3

C:Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like

Query Match 51.1%; Score 48; DB 2; Length 3704;
 Best Local Similarity 77.8%; Pred. No. 50;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 CVKCKPNFY 9
 | : | | | | | | | |
 Db 685 CDCKCPNPFY 693

RESULT 5

S37902
 hypothetical protein YKL077w - yeast (*Saccharomyces cerevisiae*)

C:Species: *Saccharomyces cerevisiae*

C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 04-Mar-2000

C:Accession: S37902

R:Pohl, T.M.; Pohl, F.M.
 submitted to the Protein Sequence Database, March 1994

A:Reference number: S37897

A:Accession: S37902

A:Molecule type: DNA

A:Residues: 1-392 <POH>

A:Cross-references: EMBL:Z28077; NID:9486109; PID:9486110; GSPDB:GN00011; MIPS:YKL077

A:Experimental source: strain S288C

C:Genetics:

A:Gene: MIPS:YKL077w

A:Map position: 11L

C:Superfamily: *Saccharomyces cerevisiae* hypothetical protein YKL077w

Query Match 50.0%; Score 47; DB 2; Length 392;
 Best Local Similarity 58.3%; Pred. No. 11;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 VKCKPNFYNGG 13
 | : | | | | | | | |
 Db 153 VRCFPEFLYFNKG 164

RESULT 6

T20465
 hypothetical protein F01D5.7 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000

C:Accession: T20465

R:Lloyd, C.
 submitted to the EMBL Data Library, November 1996

A:Reference number: Z19279

A:Accession: T20465

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-335 <NTL>

A:Cross-references: EMBL:Z81493; PTDN:CA804038.1; GSPDB:GN00020; CESP:F01D5.7

A:Experimental source: clone F01D5

C:Genetics:

A:Gene: CESP:F01D5.7

A:Map position: 2

A:Introns: 6/3; 121/3; 162/2; 194/3; 244/2; 309/2

Query Match 47.9%; Score 45; DB 2; Length 335;
 Best Local Similarity 77.8%; Pred. No. 18;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CVKCKPNFY 9
 | | | | | | | | | |
 Db 104 CVKCPNCFY 112

RESULT 7

T13711
 polyprotein - fruit fly (*Drosophila melanogaster*) retrotransposon-like element

C:Species: *Drosophila melanogaster*

C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 20-Jun-2000

C:Accession: T13711

R:Tulin, A.V.; Kogan, G.L.
 submitted to the EMBL Data Library, August 1998

A:Description: New retrotransposon of *Drosophila melanogaster*.

A:Reference number: Z17704

A:Accession: T13711

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1571 <TUL>

A:Cross-references: EMBL:AJ010298; PTDN:CA09069.1

C:Genetics:
A:Mobile element: retrotransposon-like element

Query Match 47.9%; Score 45; DB 2; Length 1571;
Best Local Similarity 77.8%; Pred. No. 68;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 CKPKNFYNG 12
|||||1:1
Db 463 CKPKNFYNG 471

RESULT 8
T12744
pseudomurein endoisopeptidase - Methanobacterium phage psIM2

C:Species: Methanobacterium phage psIM2
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 04-Mar-2000
C:Accession: T12744

R:Piester, P.; Wasserfallen, A.; Stettler, R.; Leisinger, T.
submitted to the EMBL Data Library, May 1998

A:Description: Archaeophage psIM2 complete genomic DNA.

A:Reference number: 217578

A:Accession: T12744

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-305 <PFI>

A:Cross-references: EMBL:AF065411; NID:g3249585; PID:g3249613; PIDN:AAC7067.1

A:Experimental source: Host Methanobacterium thermoautotrophicum strain Marburg

C:Genetics:

A:Gene: peip

C:Superfamily: Methanobacterium phage psIM2 pseudomurein endoisopeptidase

Query Match 46.8%; Score 44; DB 2; Length 305;
Best Local Similarity 53.8%; Pred. No. 24;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 3 KCKPKNFYNGSP 15
|||11111:
Db 186 KCKPKNFYNGSP 198

RESULT 9

T30370
hypothetical protein ORF23 - Lymantria dispar nuclear polyhedrosis virus

C:Species: Lymantria dispar nuclear polyhedrosis virus; LDMNV

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T30370

R:Kuzio, J.; Pearson, M.N.; Harwood, S.H.; Funk, C.J.; Evans, J.T.; Slavicek, J.M.; Rohl

Virolgy 253, 17-34, 1999

A:Title: Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria d

A:Reference number: Z20836; MUID:99124785

A:Accession: T30370

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-342 <KUZ>

A:Cross-references: EMBL:AF081810; PIDN:AAC70208.1

Query Match 46.8%; Score 44; DB 2; Length 342;
Best Local Similarity 54.5%; Pred. No. 27;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 CVKCKPKNFYNG 11
|:|11111
Db 271 CIRCCKSREYKN 281

RESULT 10

T15626
hypothetical protein C25H3.8 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15626

R:Johnson, D.

submitted to the EMBL Data Library, June 1995

A:Description: The sequence of C. elegans cosmid C25H3.

A:Reference number: Z18379

A:Accession: T15626

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2102 <JOH>

A:Cross-references: EMBL:U29535; NID:g868251; PID:g868260; PIDN:AAA68789.1; CESP:C25H

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:C25H3.8

A:Introns: 400/2; 464/1; 613/3; 669/3; 682/1; 927/1; 1266/3; 1403/3; 1595/1; 1625/1;

Query Match 46.8%; Score 44; DB 2; Length 2102;
Best Local Similarity 72.7%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 KCKPKNFYNG 13
|||11111:
Db 1215 KCKPKNFYNG 1225

RESULT 11

S61885
extensin precursor (clone 6Pext1.2) - wood tobacco

C:Species: Nicotiana glauca (wood tobacco)

C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 24-Nov-1999
C:Accession: S61885

R:Parmentier, Y.; Durr, A.; Marbach, J.; Hirsinger, C.; Criqui, M.C.; Fleck, J.; Jame

Plant Mol. Biol. 29, 279-292, 1995

A:Title: A novel wound-inducible extensin gene is expressed early in newly isolated p

A:Reference number: S61885; MUID:96046747

A:Accession: S61885

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-139 <PAR>

A:Cross-references: EMBL:X70343

C:Superfamily: proline-rich peptide P-B

Query Match 45.7%; Score 43; DB 2; Length 139;
Best Local Similarity 42.9%; Pred. No. 18;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 2 VKCKPKNFYNGSP 15
|||11111:
Db 22 LECKKNFYNSPP 35

RESULT 12

S74667
hypothetical protein slr1796 - Synechocystis sp. (strain PCC 6803)

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S74667

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,

O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys

S.

A:Reference number: S74322; MUID:97061201

A:Accession: S74667

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-201 <KAN>

A:Cross-references: EMBL:D90900; GB:AB001339; NID:g1651766; PIDN:BAA16819.1; PID:g165

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Superfamily: Synechocystis hypothetical protein slr1796

Query Match	45.7%;	Score 43;	DB 2;	Length 201;
Best Local Similarity	54.5%;	Pred. No. 24;		
Matches	6;	Conservative	3;	Mismatches 2;
				Indels

QY	5	KPNFYNYNGGSP	15
		: : : :	
Db	134	EPGYYYSSGVP	144

RESULT 13
H69414
conserved hypotheotical protein AF1321 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: H69414
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
J., Fliedlschmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirschner, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Ueberbach, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A:Reference number: A69250; MUID:98049343
A:Accession: H69414
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1,287 <RLE>
A:Cross-references: GB:AE001012; GB:AE000782; MID:g2669335; PIDN:AB689924.1; PID:g264922
A:Superfamily: conserved hypotheotical protein M01157

Query Match	45.7%	Score 43;	DB 1;	Length 287;
Best Local Similarity	63.6%;	Pred. NO. 32;		
Matches	7;	Conservative	0;	Mismatches 4;
			Indels	0;
			Gaps	0;

QY	1	CVKCKPNEYYN	11
Db	22	CEKCYPEFYRN	32

RESULT 14
T46383
hypothetical protein DKFZp434O1519.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46383
R:Ottenwaelder, B.; Obermayer, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23031
A:Accession: T46383
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-400 <AAA>
A:Cross-references: EMBL:AL137540
A:Experimental source: adult testis; clone DKFZp434O1519
C:Genetics:
/:Note: DKFZp434O1519.1

Query Match	45.7%;	Score 43;	DB 2;	Length 400;
Best Local Similarity	66.7%;	Pred. No. 43;		
Matches	6;	Conservative	1;	Mismatches 2;
			Indels	0;
			Gaps	0;

```
QY      1  CVKCKPNEY  9
          | : ||| | |
Db      143  CQRCKPGFY 151
```

RESULT 15
MMMSB2
laminin gamma-1 chain precursor - mouse

N: Alternate names: laminin chain B2
C: Species: Mus musculus (house mouse)
C: Date: 28-Feb-1996 #sequence_revision 30-Jun-1991 #text_change 10-Dec-1999
C: Accession: A28469; A27729; A28082; S02680; S05327; S02037; A02870; S13544; S14552
R: Sasaki, M.; Yamada, Y.
J: Biol. Chem. 262, 17111-17117, 1987
A: Title: The laminin B2 chain has a multidomain structure homologous to the B1 chain.
A: Reference number: A28469; MUID:88059118
A: Accession: A28469
A: Molecule type: mRNA
A: Residues: 1-1607 <SAS>
A: Cross-references: EMBL:J03484; MID:9198694; PIDs:AAA39405.1; PID:9293688
R: Durkin, M.E.; Barros, B.B.; Liu, S.H.; Phillips, S.L.; Chung, A.E.
Biochemistry 27, 5199-5204, 1988
A: Title: Primary structure of the mouse laminin B2 chain and comparison with laminin
A: Reference number: A27729; MUID:89000737
A: Accession: A27729
A: Molecule type: mRNA
A: Residues: 1-263, 'D', 265-336, 'C', 338-446, 'PS', 449-661, 'S', 663-885, 887-1155, 1157-1433
A: Cross-references: EMBL:J02930; MID:9198702; PIDs:AAA39408.1; PID:9293691
A: Note: the authors translated the codon TAT for residue 544 as Asp and GCG for resid
J: Ogawa, K.; Burdello, P.D.; Sasaki, M.; Yamada, Y.
J: Biol. Chem. 263, 8384-8389, 1988
A: Title: The laminin B2 chain promoter contains unique repeat sequences and is active
A: Reference number: A28082; MUID:88228071
A: Accession: A28082
A: Molecule type: DNA
A: Residues: 1-215, 'A', 217-239 <OGA>
A: Cross-references: EMBL:J03749; MID:9198704; PIDs:AAA39409.1; PID:9554184
R: Fujiwara, S.; Shinkai, H.; Deutzmann, R.; Paulsson, M.; Timpl, R.
Biochem. J. 255, 453-461, 1988
A: Title: Structure and distribution of N-linked oligosaccharide chains on various dom
A: Reference number: S02678; MUID:88326259
A: Accession: S02680
A: Molecule type: protein
A: Residues: 227-238 <FUJ>
R: Hartl, L.; Oberhaeumer, I.; Deutzmann, R.
Eur. J. Biochem. 173, 629-635, 1988
A: Title: The N terminus of laminin A chain is homologous to the B chains.
A: Reference number: S00624; MUID:88225080
A: Accession: S05327
A: Molecule type: protein
A: Residues: 227-238, 387-393, 'F', 395-405, 881-912, 1022-1034 <HAR>
R: Deutzmann, R.; Huber, J.; Schmetz, K.A.; Oberhaeumer, I.; Hartl, L.
Eur. J. Biochem. 177, 35-45, 1988
A: Title: Structural study of long arm fragments of laminin. Evidence for repetitive C
A: Reference number: S01790; MUID:89030693
A: Accession: S02037
A: Molecule type: protein
A: Residues: 1362-1377, 'X', 1379-1392, 'X', 1394-1406 <DEU>
R: Barlow, D.P.; Green, N.M.; Kurkinen, M.; Hogan, B.L.M.
EMBO J. 3, 2355-2362, 1984
A: Title: Sequencing of laminin B chain cDNAs reveals C-terminal regions of coiled-coil
A: Reference number: A02870; MUID:85051302
A: Accession: A02870
A: Molecule type: mRNA
A: Residues: 1591-1474, 'K', 1476-1575, 'N', 1577-1607 <BAR>
A: Cross-references: EMBL:X05211; MID:952862; PIDs:CAA28838.1; PID:9817975
R: Paulsson, M.; Deutzmann, R.; Timpl, R.; Dalzoppo, D.; Odermatt, E.; Engel, J.
EMBO J. 4, 309-316, 1985
A: Title: Evidence for coiled-coil alpha-helical regions in the long arm of laminin,
A: Reference number: S13543; MUID:85257455
A: Accession: S13544
A: Molecule type: protein
A: Residues: 1506-1523, 'X', 1525 <PAU>
R: Olsen, D.; Nagayoshi, T.; Fazio, M.; Peltonen, J.; Jaakkola, S.; Sanborn, D.; Sasaki
Lab. Invest. 60, 772-782, 1989
A: Title: Human laminin: Cloning and sequence analysis of cDNAs encoding A, B1 and B2
A: Reference number: A34961; MUID:89280632
A: Accession: S14552
A: Molecule type: protein
A: Residues: 801-912, 1022-1034, 1364-1377, 1379-1392, 1394-1409, 1506-1525, 1593-1606 <OLS>
C: Genetics:

A:Gene: Lamb-2
 A:Map position: 1
 A:Introns: 138/1; 239/3
 C:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin
 C:Function:
 A:Description: Interact with cells and with other basement membrane proteins to promote
 C:Superfamily: laminin beta-1 chain; laminin-type EGF-like homology
 C:Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellular
 F:1-33/Domain: signal sequence #status predicted <SIG>
 F:34-1607/Product: laminin gamma-1 chain #status predicted <MAT>
 F:34-283/Domain: VI <DOM6>
 F:284-502/Domain: V <DOM5>
 F:284-337/Domain: laminin-type EGF-like homology #status atypical <LE01>
 F:340-393/Domain: laminin-type EGF-like homology <LE02>
 F:396-440/Domain: laminin-type EGF-like homology <LE03>
 F:443-490/Domain: laminin-type EGF-like homology <LE04>
 F:493-502/Domain: laminin-type EGF-like homology #status atypical <LE05>
 F:503-687/Domain: IV <DOM4>
 F:688-1032/Domain: III <DOM3>
 F:688-719/Domain: laminin-type EGF-like homology #status atypical <LE06>
 F:722-768/Domain: laminin-type EGF-like homology <LE07>
 F:771-823/Domain: laminin-type EGF-like homology <LE08>
 F:826-879/Domain: laminin-type EGF-like homology <LE09>
 F:882-930/Domain: laminin-type EGF-like homology <LE10>
 F:933-978/Domain: laminin-type EGF-like homology <LE11>
 F:981-1026/Domain: laminin-type EGF-like homology <LE12>
 F:1033-1607/Domain: II/I <DOM2>
 F:1033-1607/Region: heptad repeats
 F:38-48/Disulfide bonds: #status predicted
 F:58-132,574,648,1020,1105,1159,1173,1203,1221,1239,1437/Binding site: carbohydrate (Asn
 F:1029,1033/Disulfide bonds: interchain #status predicted
 F:1378,1393/Binding site: carbohydrate (Asn) (covalent) #status experimental
 F:1598/Disulfide bonds: interchain (to chain B1) #status experimental

Query Match 45.7% Score 43; DB 1; Length 1607;
 Best Local Similarity 40.0% Pred. No. 1,4e+02;
 Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
 QY 1 CVKCKPNEYNGSP 15
 Db 1010 CDOCEENFYNRSWP 1024

Search completed: March 6, 2001, 12:50:54
 Job time: 175 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:54:37 ; Search time 38.83 Seconds
(without alignments)
12.475 Million cell updates/sec

Title: US-09-196-161d-6
Perfect score: 94
Sequence: 1 CVCKCPNFYNGGSP 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues
Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	52.1	415	1 TNRC_MOUSE	P50284 mus musculus
2	48	51.1	3672	1 LML2_CAEEL	Q21313 caenorhabdi
3	47	50.0	392	1 YKH7_YEAST	P36081 saccharomyc
4	43	45.7	1607	1 LMG1_MOUSE	P02468 mus musculu
5	43	45.7	1609	1 LMG1_HUMAN	P11047 homo sapien
6	42	44.7	335	1 FASA_HUMAN	P25445 homo sapien
7	42	44.7	803	1 UBPE_YEAST	P38237 saccharomyc
8	42	44.7	2444	1 NTCL_HUMAN	P46531 homo sapien
9	42	44.7	2531	1 NTCL_MOUSE	Q01705 mus musculu
10	42	44.7	3075	1 LMA1_HUMAN	Q07008 rattus norv
11	42	44.7	3084	1 LMA1_MOUSE	P25391 homo sapien
12	42	44.7	3712	1 LMA1_MOUSE	Q00174 dtrosophila
13	42	44.7	1291	1 YC81_CAEEL	Q19981 caenorhabdi
14	41.5	44.1	65	1 IVB3_VIPPA	P00992 pongo pygma
15	41	43.6	273	1 ZN80_PONPY	P36941 homo sapien
16	41	43.6	435	1 TNRC_HUMAN	P15157 rattus norv
17	41	43.6	455	1 P2X5_RAT	Q03600 caenorhabdi
18	41	43.6	1139	1 YMB3_CAEEL	P24348 caenorhabdi
19	41	43.6	1323	1 LMG1_CAEEL	P15215 dtrosophila
20	41	43.6	1639	1 LMG1_MOUSE	P21783 xenopus lae
21	41	43.6	2524	1 LMA5_MOUSE	O61001 mus musculu
22	41	43.6	3635	1 PGBM_MOUSE	O05793 mus musculu
23	41	43.6	3707	1 SMTD_HUMAN	P21589 homo sapien
24	41	43.1	574	1 IYB1_NAJNA	P20228 naja naja (
25	40.5	42.6	57	1 ZN80_GORGO	P51503 gorilla gor
26	40	42.6	273	1 ZN80_HUMAN	P51506 pan troglod
27	40	42.6	273	1 ZN80_PANTR	P51505 macaca mula
28	40	42.6	293	1 FASA_MOUSE	P25446 mus musculu
29	40	42.6	327	1 AMBP_MOUSE	Q07436 mus musculu
30	40	42.6	349	1 AMBP_MOUSE	O64240 rattus norv
31	40	42.6	349	1 AMBP_MOUSE	
32	40	42.6	349	1 AMBP_MOUSE	
33	40	42.6	349	1 AMBP_MOUSE	

34	40	42.6	369	1 TGT_THEMA	Q9x1p7 thermotoga
35	40	42.6	378	1 ASPG_CAEEL	Q21697 caenorhabdi
36	40	42.6	467	1 AFCL_ARATH	P51566 arabidopsis
37	40	42.6	754	1 PAIY_PINTA	P52777 pinus taeda
38	39.5	42.0	3110	1 LMA2_HUMAN	P24043 homo sapien
39	39	41.5	138	1 RBS_PORAE	Q09125 porphyridiu
40	39	41.5	211	1 UL92_HSV7J	P52471 herpes simp
41	39	41.5	277	1 MCRA_ECOCI	P24200 escherichia
42	39	41.5	297	1 APOH_RAT	P26644 rattus norv
43	39	41.5	325	1 VPI2_SFVKA	P25943 shope fibro
44	39	41.5	426	1 YXK4_YEAST	P33857 saccharomyc
45	39	41.5	551	1 YK27_YEAST	P36113 saccharomyc

ALIGNMENTS

RESULT 1
ID TNRC_MOUSE STANDARD: PRT; 415 AA.
AC P50284;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE LYMPHOTOXIN-BETA RECEPTOR PRECURSOR.
GN LTRR OR TNRC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN-CVB: TRISUE-LUNG;
RX MEDLINE=96072804; PubMed=7594541;
RA Browning J.L., Ware C.F.;
RA "Mouse lymphotoxin-beta receptor. Molecular genetics, ligand binding,
RT and expression.";
RT J. Immunol. 155:5280-5288(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96163885; PubMed=8586432;
RA Nakamura T., Tashiro K., Nazarea M., Nakano T., Sasayama S.,
RA Honjo T.;
RT "The murine lymphotoxin-beta receptor cDNA: isolation by the signal
RT sequence trap and chromosomal mapping.";
RL Genomics 30:312-319(1995).
CC - FUNCTION: RECEPTOR FOR THE LYMPHOTOXIN-BETA. POSSIBLE FUNCTION IN
CC IMMUNE DEVELOPMENT.
CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC - SIMILARITY: CONTAINS A LA-NGFR/TNFR-TYPE CYSTEINE-RICH REGION.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

CC EMBL: U29173; AA68964.1; -
DR EMBL: L38423; AAB00846.1; -
DR EMBL: U30798; AAB81334.1; -
DR HSSP: P25942; ICDF.
DR MGD: MGI:104875; LTRR.
DR INTERPRO: IPR001368; -
DR PFAM: PF00020; TNFR_C6; 3.
DR PROSITE: PS00652; TNFR_NGFR_1; 2.
DR PROSITE: PS00505; TNFR_NGFR_2; 3.
KW Receptor; Transmembrane; Glycoprotein; Repeat; Signal.
FT SIGNAL 1 30
FT CHAIN 31 415 LYMPHOTOXIN-BETA RECEPTOR.
FT DOMAIN 31 223 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 224 244 POTENTIAL.

```

FT DOMAIN 245 415 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 42 213 4 X TNFR-CYS.
FT REPEAT 42 81 TNFR-CYS 1.
FT REPEAT 82 124 TNFR-CYS 2.
FT REPEAT 125 170 TNFR-CYS 3.
FT REPEAT 171 213 TNFR-CYS 4.
FT DISULFID 43 58 BY SIMILARITY.
FT DISULFID 59 72 BY SIMILARITY.
FT DISULFID 62 80 BY SIMILARITY.
FT DISULFID 83 98 BY SIMILARITY.
FT DISULFID 101 116 BY SIMILARITY.
FT DISULFID 104 124 BY SIMILARITY.
FT DISULFID 126 132 BY SIMILARITY.
FT DISULFID 139 150 BY SIMILARITY.
FT DISULFID 142 169 BY SIMILARITY.
FT DISULFID 172 187 BY SIMILARITY.
FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 415 AA; 44956 MW; 29B326A566AEF661 CRC64;

Query Match 52.1%; Score 49; DB 1; Length 415;
Best Local Similarity 53.3%; Pred. No. 2;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 CVCKKPNFYNGSP 15
   1111 : 111
Db 169 CVPCKPGHFQNTSP 183

RESULT 2
LML2_CAEEL
ID LML2_CAEEL STANDARD; PRT; 3672 AA.
AC Q21313;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE LAMININ-LIKE PROTEIN K08C7.3 PRECURSOR.
GN K08C7.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Berks M.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -1- SIMILARITY: CONTAINS 1 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z70286; CAA94293.1; -
DR HSSP: P02468; JKLO
DR WORMPEP; K08C7.3; CE06136.
DR INTERPRO: IPR000034; -
DR INTERPRO: IPR000561; -
DR INTERPRO: IPR001791; -
DR INTERPRO: IPR001866; -
DR INTERPRO: IPR002049; -
DR PFAM: PF00052; laminin_B; 1.
DR PFAM: PF00053; laminin_EGF; 21.
DR PFAM: PF00054; laminin_G; 5.
DR PFAM: PF00055; laminin_Nterm; 1.
DR PRINTS: PR00011; EGF_LAMININ.
DR PROSITE: PS00022; EGF_1; 19.

```

```

DR PROSITE: PS01186; EGF_2; 4.
DR PROSITE: PS01248; LAMININ_type_EGF; 21.
KW Hypothetical protein; Laminin EGF-like domain; Signal; Repeat.
FT SIGNAL 1 27
FT CHAIN 28 3672 LAMININ-LIKE PROTEIN K08C7.3.
FT DOMAIN 28 297 LAMININ N-TERMINAL (DOMAIN VI).
FT DOMAIN 298 356 LAMININ EGF-LIKE 1.
FT DOMAIN 357 426 LAMININ EGF-LIKE 2.
FT DOMAIN 427 471 LAMININ EGF-LIKE 3.
FT DOMAIN 472 518 LAMININ EGF-LIKE 4.
FT DOMAIN 519 563 LAMININ EGF-LIKE 5.
FT DOMAIN 564 609 LAMININ EGF-LIKE 6.
FT DOMAIN 610 655 LAMININ EGF-LIKE 7.
FT DOMAIN 656 700 LAMININ EGF-LIKE 8.
FT DOMAIN 701 755 LAMININ EGF-LIKE 9.
FT DOMAIN 756 808 LAMININ EGF-LIKE 10.
FT DOMAIN 809 839 LAMININ EGF-LIKE 11 (INCOMPLETE).
FT DOMAIN 1415 1460 LAMININ EGF-LIKE 12.
FT DOMAIN 1461 1505 LAMININ EGF-LIKE 13.
FT DOMAIN 1506 1553 LAMININ EGF-LIKE 14.
FT DOMAIN 1554 1604 LAMININ EGF-LIKE 15.
FT DOMAIN 1605 1614 LAMININ EGF-LIKE 16 (N-TERMINAL).
FT DOMAIN 1615 1796 LAMININ EGF-LIKE 17.
FT DOMAIN 1797 1829 LAMININ EGF-LIKE 18.
FT DOMAIN 1830 1879 LAMININ EGF-LIKE 19.
FT DOMAIN 1880 1936 LAMININ EGF-LIKE 20.
FT DOMAIN 1937 1989 LAMININ EGF-LIKE 21.
FT DOMAIN 1990 2036 LAMININ EGF-LIKE 22.
FT DOMAIN 2037 2083 LAMININ EGF-LIKE 23.
FT DOMAIN 2084 2131 LAMININ EGF-LIKE 24.
FT DISULFID 298 307 BY SIMILARITY.
FT DISULFID 300 320 BY SIMILARITY.
FT DISULFID 322 331 BY SIMILARITY.
FT DISULFID 334 354 BY SIMILARITY.
FT DISULFID 357 366 BY SIMILARITY.
FT DISULFID 359 391 BY SIMILARITY.
FT DISULFID 394 403 BY SIMILARITY.
FT DISULFID 406 424 BY SIMILARITY.
FT DISULFID 427 438 BY SIMILARITY.
FT DISULFID 429 445 BY SIMILARITY.
FT DISULFID 447 456 BY SIMILARITY.
FT DISULFID 459 469 BY SIMILARITY.
FT DISULFID 472 484 BY SIMILARITY.
FT DISULFID 474 491 BY SIMILARITY.
FT DISULFID 493 502 BY SIMILARITY.
FT DISULFID 505 516 BY SIMILARITY.
FT DISULFID 519 531 BY SIMILARITY.
FT DISULFID 521 538 BY SIMILARITY.
FT DISULFID 540 549 BY SIMILARITY.
FT DISULFID 552 561 BY SIMILARITY.
FT DISULFID 564 576 BY SIMILARITY.
FT DISULFID 566 583 BY SIMILARITY.
FT DISULFID 585 594 BY SIMILARITY.
FT DISULFID 597 607 BY SIMILARITY.
FT DISULFID 610 622 BY SIMILARITY.
FT DISULFID 612 629 BY SIMILARITY.
FT DISULFID 631 640 BY SIMILARITY.
FT DISULFID 643 653 BY SIMILARITY.
FT DISULFID 656 668 BY SIMILARITY.
FT DISULFID 658 674 BY SIMILARITY.
FT DISULFID 676 685 BY SIMILARITY.
FT DISULFID 688 698 BY SIMILARITY.
FT DISULFID 701 715 BY SIMILARITY.
FT DISULFID 724 724 BY SIMILARITY.
FT DISULFID 726 735 BY SIMILARITY.
FT DISULFID 738 753 BY SIMILARITY.
FT DISULFID 756 770 BY SIMILARITY.
FT DISULFID 770 777 BY SIMILARITY.
FT DISULFID 779 788 BY SIMILARITY.
FT DISULFID 791 806 BY SIMILARITY.
FT DISULFID 1415 1427 BY SIMILARITY.
FT DISULFID 1417 1434 BY SIMILARITY.
FT DISULFID 1436 1445 BY SIMILARITY.

```

```

FT DISULFID 1448 1458 BY SIMILARITY.
FT DISULFID 1461 1469 BY SIMILARITY.
FT DISULFID 1463 1476 BY SIMILARITY.
FT DISULFID 1478 1487 BY SIMILARITY.
FT DISULFID 1490 1503 BY SIMILARITY.
FT DISULFID 1506 1520 BY SIMILARITY.
FT DISULFID 1508 1527 BY SIMILARITY.
FT DISULFID 1529 1538 BY SIMILARITY.
FT DISULFID 1541 1551 BY SIMILARITY.
FT DISULFID 1554 1566 BY SIMILARITY.
FT DISULFID 1573 1584 BY SIMILARITY.
FT DISULFID 1575 1584 BY SIMILARITY.
FT DISULFID 1587 1602 BY SIMILARITY.
FT DISULFID 1830 1839 BY SIMILARITY.
FT DISULFID 1832 1846 BY SIMILARITY.
FT DISULFID 1849 1858 BY SIMILARITY.
FT DISULFID 1861 1877 BY SIMILARITY.
FT DISULFID 1880 1894 BY SIMILARITY.
FT DISULFID 1882 1905 BY SIMILARITY.
FT DISULFID 1907 1916 BY SIMILARITY.
FT DISULFID 1919 1934 BY SIMILARITY.
FT DISULFID 1937 1951 BY SIMILARITY.
FT DISULFID 1939 1958 BY SIMILARITY.
FT DISULFID 1961 1970 BY SIMILARITY.
FT DISULFID 1973 1987 BY SIMILARITY.
FT DISULFID 1990 2000 BY SIMILARITY.
FT DISULFID 1992 2007 BY SIMILARITY.
FT DISULFID 2009 2018 BY SIMILARITY.
FT DISULFID 2021 2034 BY SIMILARITY.
FT DISULFID 2037 2048 BY SIMILARITY.
FT DISULFID 2039 2055 BY SIMILARITY.
FT DISULFID 2057 2066 BY SIMILARITY.
FT DISULFID 2069 2081 BY SIMILARITY.
FT DISULFID 2084 2096 BY SIMILARITY.
FT DISULFID 2103 2103 BY SIMILARITY.
FT DISULFID 2105 2114 BY SIMILARITY.
FT DISULFID 2117 2129 BY SIMILARITY.
FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 511 511 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 530 530 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 634 634 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 761 761 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1014 1014 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1341 1341 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1705 1705 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1756 1756 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1868 1868 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1944 1944 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1986 1986 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2002 2002 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2159 2159 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2207 2207 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2231 2231 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2235 2235 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2401 2401 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2421 2421 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2487 2487 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2821 2821 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3087 3087 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3242 3242 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3541 3541 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3672 AA; 404223 MW; 28E262DB5FF14BFA CRC64;

```

Query Match 51.1%; Score 48; DB 1; Length 3672;
 Best Local Similarity 77.8%; Pred. No. 22;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 CVCKPFFY 9

```

DB 685 CDCKPFFY 693
RESULT 3
YKH7_YEAST STANDARD; PRT: 392 AA.
AC P36081;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHETICAL 46.0 KDA PROTEIN IN SMK1-MUD2 INTERGENIC REGION.
GN YKL077W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RA Pohl T.M., Pohl F.M.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Z28077; CAAB1914.1; -
DR PIR: S37902; S37902.
DR SGD: S0001560; YKL077W.
KW Hypothetical protein.
SQ SEQUENCE 392 AA; 46036 MW; 8EEC795903D5D540 CRC64;

```

Query Match 50.0%; Score 47; DB 1; Length 392;
 Best Local Similarity 58.3%; Pred. No. 3.9;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

OY 2 CVCKPFFYNGC 13
DB 153 VRCTPVLFFKNG 164
RESULT 4
LMG1_MOUSE STANDARD; PRT: 1607 AA.
AC P02468;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JUL-1986 (Rel. 11, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE LAMININ GAMMA-1 CHAIN PRECURSOR (LAMININ B2 CHAIN).
GN LAMC1 OR LAMC-1 OR LAMB-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-86059118; PubMed-3680290;
RA Sasaki M., Yamada Y.;
RT "The laminin B2 chain has a multidomain structure homologous to the
RT B1 chain."
RL J. Biol. Chem. 262:17111-17117(1987).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE-89000737; PubMed-3167041;
RA Durkin M.E., Barros B.B., Liu S.-H., Phillips S.L., Chung A.E.;
RT "Primary structure of the mouse laminin B2 chain and comparison with
RL laminin B1."
RN Biochemistry 27:5198-5204(1988).
RN [3]
RP SEQUENCE OF 1-239 FROM N.A.

```



```

FT CAROHD 1173 1173 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 1203 1203 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 1221 1221 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 1239 1239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 1378 1378 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 1393 1393 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 1437 1437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 216 216 G -> A (IN REF. 3).
FT CONFLICT 260 260 E -> D (IN REF. 2).
FT CONFLICT 337 337 S -> C (IN REF. 2).
FT CONFLICT 447 448 LR -> PS (IN REF. 2).
FT CONFLICT 544 544 D -> Y (IN REF. 2).
FT CONFLICT 662 662 T -> S (IN REF. 2).
FT CONFLICT 886 886 MISSING (IN REF. 2).
FT CONFLICT 1158 1158 MISSING (IN REF. 2).
FT CONFLICT 1434 1434 V -> A (IN REF. 2).
FT CONFLICT 1475 1475 R -> K (IN REF. 4).
FT CONFLICT 1576 1576 D -> N (IN REF. 4).
SO SEQUENCE 1607 AA; 177297 MW; 8187808E4869F242 CRC64;

```

```

Query Match 45.7%; Score 43; DB 1; Length 1607;
Best Local Similarity 40.0%; Pred. NO. 58;
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 CVCKRPFYNGSP 15
DB 1010 CDQCEYFNYSWP 1024

```

```

RESULT 5
LMG1_HUMAN STANDARD: PRT: 1609 AA.
AC P11047:
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE LAMININ GAMMA-1 CHAIN PRECURSOR (LAMININ B2 CHAIN).
GN LAMC1 OR LAMB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=91093128; PubMed=1985895;
RX Kallunki T., Ikonen J., Chow L.T., Kallunki P., Tryggvason K.;
RT "Structure of the human laminin B2 chain gene reveals extensive
RT divergence from the laminin B1 chain gene."
RL J. Biol. Chem. 266:221-228(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88198245; PubMed=3360804;
RA Pikkarainen T., Kallunki T., Tryggvason K.;
RT "Human laminin B2 chain. Comparison of the complete amino acid
RT sequence with the B1 chain reveals variability in sequence homology
RT between different structural domains."
RL J. Biol. Chem. 263:6751-6758(1988).
RN [3]
RP SEQUENCE OF 1393-1609 FROM N.A.
RX MEDLINE=89169663; PubMed=3234037;
RA Fukushima Y., Pikkarainen T., Kallunki T., Eddy R.L., Byers M.G.,
RA Haley L.L., Henry W.M., Tryggvason K., Shows T.B.;
RT "Isolation of a human laminin B2 (LAMB2) cDNA clone and assignment of
RT the gene to chromosome region 19q25-->q31."
RL Cytogenet. Cell Genet. 48:137-141(1988).
RN [4]
RP SEQUENCE OF 1282-1609 FROM N.A.
RX TISSUE-ENDOTHELIAL CELLS;
RX MEDLINE=92216129; PubMed=1806043;
RA Santos C.L.S., Sabpaga J., Brentani R.;
RT "Differences in human laminin B2 sequences."
RL DNA Seq. 1:275-277(1991).
CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ

```

```

CC IS THOUGHT TO MEDiate THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
CC THE GAMMA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (HNS LAMININ),
CC LAMININ-2 (MEROSIN), LAMININ-3 (S-LAMININ), LAMININ-4 (S-MEROSIN),
CC LAMININ-6 (K-LAMININ), AND LAMININ-7 (KS-LAMININ).
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR
CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
CC COMPONENT).
CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -1- SIMILARITY: DOMAINS VI AND IV ARE GLOBULAR.
CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -1- SIMILARITY: CONTAINS 1 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M55210; AAAS9492.1; -
DR EMBL: M55217; AAAS9492.1; JOINED.
DR EMBL: M55201; AAAS9492.1; JOINED.
DR EMBL: M55211; AAAS9492.1; JOINED.
DR EMBL: M55212; AAAS9492.1; JOINED.
DR EMBL: M55213; AAAS9492.1; JOINED.
DR EMBL: M55214; AAAS9492.1; JOINED.
DR EMBL: M55215; AAAS9492.1; JOINED.
DR EMBL: M55216; AAAS9492.1; JOINED.
DR EMBL: M55192; AAAS9492.1; JOINED.
DR EMBL: M55193; AAAS9492.1; JOINED.
DR EMBL: M55194; AAAS9492.1; JOINED.
DR EMBL: M55195; AAAS9492.1; JOINED.
DR EMBL: M55196; AAAS9492.1; JOINED.
DR EMBL: M55197; AAAS9492.1; JOINED.
DR EMBL: M55198; AAAS9492.1; JOINED.
DR EMBL: M55199; AAAS9492.1; JOINED.
DR EMBL: M55200; AAAS9492.1; JOINED.
DR EMBL: M55202; AAAS9492.1; JOINED.
DR EMBL: M55203; AAAS9492.1; JOINED.
DR EMBL: M55204; AAAS9492.1; JOINED.
DR EMBL: M55205; AAAS9492.1; JOINED.
DR EMBL: M55206; AAAS9492.1; JOINED.
DR EMBL: M55207; AAAS9492.1; JOINED.
DR EMBL: M55208; AAAS9492.1; JOINED.
DR EMBL: M55209; AAAS9492.1; JOINED.
DR EMBL: M55209; AAAS9492.1; JOINED.
DR EMBL: M27654; AAAS9488.1; -
DR EMBL: X13939; CAA32122.1; -
DR PIR: S13548; MMBH2.
DR HSSP: P02468; TITLE.
DR MIM: 150290; -
DR INTERPRO: IPR000034; -
DR INTERPRO: IPR000561; -
DR INTERPRO: IPR001886; -
DR INTERPRO: IPR002049; -
DR PFM: PF00052; laminin_B; 1.
DR PFM: PF00053; laminin_EGF; 10.
DR PFM: PF00055; laminin_Nterm; 1.
DR PRINTS: PR00011; EGF_LAMININ.
DR PROSITE: PS00022; EGF_1; 2.
DR PROSITE: PS01186; EGF_2; 2.
DR PROSITE: PS01248; LAMININ_TYPE_EGF; 11.
DR Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.

```

```

FT SIGNAL 1 33
FT CHAIN 34 1609 LAMININ GAMMA-1 CHAIN.
FT DOMAIN 34 285 LAMININ N-TERMINAL (DOMAIN VI).
FT 286 504 4.5 X LAMININ EGF-LIKE REPEATS (DOMAIN
V).
FT DOMAIN 286 341 LAMININ EGF-LIKE 1.
FT 342 397 LAMININ EGF-LIKE 1.
FT DOMAIN 342 397 LAMININ EGF-LIKE 2.
FT 398 444 LAMININ EGF-LIKE 3.
FT 445 494 LAMININ EGF-LIKE 4.
FT 495 504 LAMININ EGF-LIKE 5 (N-TERMINAL).
FT 505 689 LAMININ DOMAIN IV.
FT 690 1029 6.5 X LAMININ EGF-LIKE REPEATS (DOMAIN
III).
FT 690 723 LAMININ EGF-LIKE 5 (C-TERMINAL).
FT 724 772 LAMININ EGF-LIKE 6.
FT 773 827 LAMININ EGF-LIKE 7.
FT 828 883 LAMININ EGF-LIKE 8.
FT 884 934 LAMININ EGF-LIKE 9.
FT 935 982 LAMININ EGF-LIKE 10.
FT 1030 1609 LAMININ EGF-LIKE 11.
FT 1038 1609 LAMININ II AND I.
FT 1609 295 COILED COIL (POTENTIAL).
FT 286 305 BY SIMILARITY.
FT 307 316 BY SIMILARITY.
FT 319 339 BY SIMILARITY.
FT 342 351 BY SIMILARITY.
FT 344 367 BY SIMILARITY.
FT 370 379 BY SIMILARITY.
FT 382 395 BY SIMILARITY.
FT 398 410 BY SIMILARITY.
FT 400 416 BY SIMILARITY.
FT 418 427 BY SIMILARITY.
FT 430 442 BY SIMILARITY.
FT 445 456 BY SIMILARITY.
FT 447 463 BY SIMILARITY.
FT 465 474 BY SIMILARITY.
FT 477 492 BY SIMILARITY.
FT 724 733 BY SIMILARITY.
FT 726 740 BY SIMILARITY.
FT 742 751 BY SIMILARITY.
FT 754 770 BY SIMILARITY.
FT 773 781 BY SIMILARITY.
FT 775 792 BY SIMILARITY.
FT 795 804 BY SIMILARITY.
FT 807 825 BY SIMILARITY.
FT 828 842 BY SIMILARITY.
FT 830 849 BY SIMILARITY.
FT 852 861 BY SIMILARITY.
FT 864 881 BY SIMILARITY.
FT 884 898 BY SIMILARITY.
FT 898 905 BY SIMILARITY.
FT 905 916 BY SIMILARITY.
FT 919 932 BY SIMILARITY.
FT 935 947 BY SIMILARITY.
FT 937 954 BY SIMILARITY.
FT 956 965 BY SIMILARITY.
FT 968 980 BY SIMILARITY.
FT 983 995 BY SIMILARITY.
FT 995 1001 BY SIMILARITY.
FT 1003 1012 BY SIMILARITY.
FT 1015 1028 BY SIMILARITY.
FT 1031 1031 INTERCHAIN (PROBABLE).
FT 1034 1034 INTERCHAIN (PROBABLE).
FT 1034 1034 INTERCHAIN (PROBABLE).
FT 1600 1600 INTERCHAIN (PROBABLE).
FT 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 650 650 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1022 1022 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1107 1107 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1161 1161 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT 1175 1175 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT CARBOHYD 1205 1205 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1223 1223 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1241 1241 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1380 1380 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1395 1395 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1439 1439 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT 212 212 F -> I (IN REF. 2).
SQ SEQUENCE 1609 AA: 177606 MW: B098F20FC97293B CRC64:

Query Match 45.7%; Score 43; DB 1; Length 1609;
Best Local Similarity 40.0%; Pred. NO. 58;
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Oy 1 CVKCKPNFYNGGSP 15
Db 1012 CDCCENFYNRSP 1026

RESULT 6
FASA_HUMAN STANDARD; PRT; 335 AA.
ID FASA_HUMAN
AC P25445.
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE FASL RECEPTOR PRECURSOR (APOPTOSIS-MEDIATING SURFACE ANTIGEN FAS)
DE (APO-1 ANTIGEN) (CD95 ANTIGEN).
DE TNFRSF6 OR APT1 OR FAS OR FASL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
[1]
SEQUENCE FROM N.A.
RX MEDLINE=91309137; PubMed=1713127;
RA Itoh N., Yonehara S., Ishii A., Yonehara M., Mizushima S.I.,
RA Sameshima M., Hase A., Seto Y., Nagata S.;
RT "The polypeptide encoded by the cDNA for human cell surface antigen
RL Fas can mediate apoptosis.";
RL Cell 66:233-243(1991).
[2]
SEQUENCE FROM N.A. AND SEQUENCE OF 226-240; 269-291 AND 321-335.
RX MEDLINE=92268122; PubMed=1375228;
RA Oehm A., Behrmann I., Falk W., Pawlita M., Maier G., Klas C.,
RA Li-Weber M., Richards S., Dhein J., Trauth B.C., Ponattingl H.,
RA Krammer P.H.;
RT "Purification and molecular cloning of the APO-1 cell surface
RT antigen, a member of the tumor necrosis factor/nerve growth factor
RT receptor superfamily. Sequence identity with the Fas antigen.";
RL J. Biol. Chem. 267:10709-10715(1992).
[3]
STRUCTURE BY NMR OF 218-335.
RX MEDLINE=97122332; PubMed=8697952;
RA Huang B., Eberstadt M., Olejniczak E.T., Meadows R.P., Fesik S.W.;
RT "NMR structure and mutagenesis of the Fas (APO-1/CD95) death domain.";
RL Nature 384:638-641(1996).
-1- FUNCTION: RECEPTOR FOR A CYTOKINE LIGAND KNOWN AS FASL. THE
ADAPTOR MOLECULE FADD RECRUITS CASPASE-8 TO THE ACTIVATED
RECEPTOR. THE RESULTING AGGREGATE CALLED THE DEATH-INDUCING
SIGNALING COMPLEX (DISC) PERFORMS CASPASE-8 PROTEOLYTIC
ACTIVATION. ACTIVE CASPASE-8 INITIATES THE SUBSEQUENT CASCADE OF
CASPASES (ASPARTATE-SPECIFIC CYSTEINE PROTEASES) MEDIATING
APOPTOSIS. FAS-MEDIATED APOPTOSIS MAY HAVE A ROLE IN THE
INDUCTION OF PERIPHERAL TOLERANCE, IN THE ANTIGEN-STIMULATED
SUICIDE OF MATURE T-CELLS, OR BOTH.
-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
-1- DOMAIN: CONTAINS A DEATH DOMAIN INVOLVED IN THE BINDING OF FADD,
AND MAYBE TO OTHER CYTOSOLIC ADAPTOR PROTEINS.
-1- SIMILARITY: CONTAINS A LA-NGER/TNFR-TYPE CYSTEINE-RICH REGION.
-1- SIMILARITY: CONTAINS 1 DEATH DOMAIN.
-1- DATABASE: NAME=PROW; NOTE=CD guide CD95 entry;
WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd95.htm".

```

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M67454; AA63174.1; -
 DR EMBL: X63717; CAA45250.1; -
 DR PIR: A40036; A40036.
 DR PIR: S24543; S24543.
 DR PDB: 1DDE; 12-NOV-97.
 DR MIM: 134637; -
 DR INTERPRO: IPR000488; -
 DR INTERPRO: IPR001368; -
 DR PFAM: PF00620; TNFR_C6; 2.
 DR PFAM: PF00531; death; 1.
 DR PROSITE: PS00652; TNFR_NGFR_1; 2.
 DR PROSITE: PS00550; TNFR_NGFR_2; 2.
 DR PROSITE: PS50017; DEATH_DOMAIN; 1.
 DR Apoptosis; Receptor; Glycoprotein; Transmembrane; Repeat; Signal;
 KW 3D-structure.
 FT SIGNAL 1 16 POTENTIAL.
 FT CHAIN 17 335 FAST RECEPTOR.
 FT DOMAIN 17 173 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 174 190 POTENTIAL.
 FT DOMAIN 191 335 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 47 166 3 X TNFR-CYS.
 FT REPEAT 83 83 TNFR-CYS 1.
 FT REPEAT 84 127 TNFR-CYS 2.
 FT REPEAT 128 166 TNFR-CYS 3.
 FT DOMAIN 230 314 DEATH DOMAIN.
 FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 335 AA; 37732 MW; 013994253511410 CRC64;
 Query Match 44.7%; Score 42; DB 1; Length 335;
 Best Local Similarity 66.7%; Pred. No. 19;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 3 KCKPNEYNN 11
 Db 128 KCKPNEPCN 136
 RESULT 7
 UBPE_YEAST STANDARD: PRT; 803 AA.
 AC P38237;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE UBQUITIN CARBOXYL-TERMINAL HYDROLASE 14 (EC 3.1.2.15) (UBIQUITIN
 DE THIOLESTERASE 14) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 14)
 DE (DUBIQUITININATING ENZYME 14).
 GN UBPI4 OR YBR058C OR YBR0515.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 CC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=95321020; PubMed=7597852;
 RA Alljovic G., Pohl T.M.;
 RT "Sequence and analysis of 24 kb on chromosome II of Saccharomyces
 RT cerevisiae".
 RL Yeast 11:475-479(1995).
 CC -1- CATALYTIC ACTIVITY: UBIQUITIN C-TERMINAL THIOLESTER + H(2)O =
 CC UBIQUITIN + A THIOL.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C19; ALSO KNOWN AS
 CC FAMILY 2 OF UBIQUITIN CARBOXYL-TERMINAL HYDROLASES.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: Z35927; CAA85001.1; -
 DR EMBL: Z46260; CAA86402.1; -
 DR PIR: S45916; S45916.
 DR MEROPS: C19.001; -
 DR SGD: S0000262; UBP14.
 DR INTERPRO: IPR000449; -
 DR INTERPRO: IPR001394; -
 DR PFAM: PF00627; UBA; 2.
 DR PFAM: PF00442; UCH-1; 1.
 DR PFAM: PF00443; UCH-2; 1.
 DR PROSITE: PS00972; UCH_2-1; 1.
 DR PROSITE: PS00973; UCH_2-2; 1.
 DR PROSITE: PS00973; UCH_2-3; 1.
 DR Ubiquitin conjugation; Hydrolase; Thiol protease; Multigene family.
 KW Ubiquitin conjugation; Hydrolase; Thiol protease; Multigene family.
 FT ACT SITE 354 354 BY SIMILARITY.
 FT ACT SITE 750 750 BY SIMILARITY.
 FT ACT SITE 759 759 BY SIMILARITY.
 SQ SEQUENCE 803 AA; 91110 MW; 8916B60E385C39A CRC64;
 Query Match 44.7%; Score 42; DB 1; Length 803;
 Best Local Similarity 50.0%; Pred. No. 43;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 2 VKCKPNEYNGSP 15
 Db 402 MKCEPELYPNGIKP 415
 RESULT 8
 NTCL_HUMAN STANDARD: PRT; 2444 AA.
 AC P46531;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1 PRECURSOR (TRANSLLOCATION-
 DE ASSOCIATED NOTCH PROTEIN TAN-1) (FRAGMENT).
 GN NOTCH1 OR TAN1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91347367; PubMed=1831692;
 RA Ellisen L.W., Bird J., West D.C., Soreng A.L., Reynolds T.C.,
 RA Smith S.D., Sklar J.;
 RT "TAN-1, the human homolog of the Drosophila notch gene, is broken by
 RT chromosomal translocations in T lymphoblastic neoplasms".
 RL Cell 66:649-661(1991).
 CC -1- FUNCTION: MAY BE IMPORTANT FOR NORMAL LYMPHOCYTE FUNCTION. IN
 CC ALTERED FORM, MAY CONTRIBUTE TO TRANSFORMATION OR PROGRESSION
 CC IN SOME T-CELL NEOPLASMS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: IN FETAL TISSUES MOST ABUNDANT IN SPLEEN,
 CC BRAIN STEM AND LUNG. ALSO PRESENT IN MOST ADULT TISSUES WHERE IT
 CC IS FOUND MAINLY IN LYMPHOID TISSUES.
 CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
 CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M73980; AAA60614.1; -
DR HSSP; P00740; 11XA.
DR MIM; 190198; -
DR INTERPRO; IPR000152; -
DR INTERPRO; IPR0000561; -
DR INTERPRO; IPR000800; -
DR INTERPRO; IPR001881; -
DR INTERPRO; IPR002110; -
DR PFAM; PF00008; EGF; 36.
DR PFAM; PF00023; ank; 6.
DR PFAM; PF00066; notch; 3.
DR PROSITE; PSS0088; ANK_REPEAT; 4.
DR PROSITE; PSS0297; ANK_REPEAT_REGION; 1.
DR PROSITE; PSS0010; ASX_HYDROXYL; 20.
DR PROSITE; PSS0022; EGF_1; 34.
DR PROSITE; PSS0186; EGF_2; 26.
DR PROSITE; PSS0187; EGF_CA; 18.
KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
Transmembrane; Signal; Glycoprotein.
FT SIGNAL 1 18
FT CHAIN 19 >2444 POTENTIAL.
FT DOMAIN 19 1736 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1.
FT TRANSMEM 1737 1757 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 1758 >2444 POTENTIAL.
FT DOMAIN 20 58 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 59 99 EGF-LIKE 1.
FT DOMAIN 102 139 EGF-LIKE 2.
FT DOMAIN 140 176 EGF-LIKE 3.
FT DOMAIN 178 216 EGF-LIKE 4.
FT DOMAIN 218 255 EGF-LIKE 5.
FT DOMAIN 257 293 EGF-LIKE 6.
FT DOMAIN 295 333 EGF-LIKE 7.
FT DOMAIN 335 371 EGF-LIKE 8.
FT DOMAIN 372 410 EGF-LIKE 9.
FT DOMAIN 412 450 EGF-LIKE 10.
FT DOMAIN 452 488 EGF-LIKE 11.
FT DOMAIN 490 526 EGF-LIKE 12.
FT DOMAIN 528 564 EGF-LIKE 13.
FT DOMAIN 566 601 EGF-LIKE 14.
FT DOMAIN 603 639 EGF-LIKE 15.
FT DOMAIN 641 676 EGF-LIKE 16.
FT DOMAIN 678 714 EGF-LIKE 17.
FT DOMAIN 716 751 EGF-LIKE 18.
FT DOMAIN 753 789 EGF-LIKE 19.
FT DOMAIN 791 827 EGF-LIKE 20.
FT DOMAIN 829 868 EGF-LIKE 21.
FT DOMAIN 870 906 EGF-LIKE 22.
FT DOMAIN 908 944 EGF-LIKE 23.
FT DOMAIN 946 982 EGF-LIKE 24.
FT DOMAIN 984 1020 EGF-LIKE 25.
FT DOMAIN 1022 1058 EGF-LIKE 26.
FT DOMAIN 1060 1096 EGF-LIKE 27.
FT DOMAIN 1098 1144 EGF-LIKE 28.
FT DOMAIN 1146 1182 EGF-LIKE 29.
FT DOMAIN 1184 1220 EGF-LIKE 30.
FT DOMAIN 1222 1266 EGF-LIKE 31.
FT DOMAIN 1268 1306 EGF-LIKE 32.
FT DOMAIN 1308 1347 EGF-LIKE 33.
FT DOMAIN 1349 1385 EGF-LIKE 34.
FT DOMAIN 1388 1427 EGF-LIKE 35.
FT DOMAIN 1446 1481 EGF-LIKE 36.
FT REPEAT 1482 1523 3 X LIN/NOTCH REPEATS.
FT REPEAT 1524 1563 LIN/NOTCH 1.
FT DOMAIN 1876 1921 LIN/NOTCH 2.
FT REPEAT 1923 1954 6 X ANK MOTIF REPEATS.
FT REPEAT ANK MOTIF 1.
FT REPEAT ANK MOTIF 2.

FT REPEAT 1956 1987 ANK MOTIF 3.
FT REPEAT 1990 2021 ANK MOTIF 4.
FT REPEAT 2023 2054 ANK MOTIF 5.
FT REPEAT 2056 2087 ANK MOTIF 6.
FT DOMAIN 1576 1579 POLY-VAL.
FT DOMAIN 1662 1665 POLY-ARG.
FT DOMAIN 1729 1732 POLY-PRO.
FT DOMAIN 1741 1744 POLY-ALA.
FT DOMAIN 1902 1905 POLY-GLU.
FT DOMAIN 2260 2263 POLY-GLY.
FT DOMAIN 2404 2407 POLY-GLN.
FT DOMAIN 2411 2418 POLY-PRO.
FT DOMAIN 24 37 BY SIMILARITY.
FT DISULFID 31 46 BY SIMILARITY.
FT DISULFID 48 57 BY SIMILARITY.
FT DISULFID 63 74 BY SIMILARITY.
FT DISULFID 68 87 BY SIMILARITY.
FT DISULFID 89 98 BY SIMILARITY.
FT DISULFID 106 117 BY SIMILARITY.
FT DISULFID 111 127 BY SIMILARITY.
FT DISULFID 129 138 BY SIMILARITY.
FT DISULFID 144 155 BY SIMILARITY.
FT DISULFID 149 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 182 195 BY SIMILARITY.
FT DISULFID 189 204 BY SIMILARITY.
FT DISULFID 206 215 BY SIMILARITY.
FT DISULFID 222 233 BY SIMILARITY.
FT DISULFID 227 243 BY SIMILARITY.
FT DISULFID 245 254 BY SIMILARITY.
FT DISULFID 261 272 BY SIMILARITY.
FT DISULFID 266 281 BY SIMILARITY.
FT DISULFID 283 292 BY SIMILARITY.
FT DISULFID 299 312 BY SIMILARITY.
FT DISULFID 306 321 BY SIMILARITY.
FT DISULFID 323 332 BY SIMILARITY.
FT DISULFID 339 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 361 370 BY SIMILARITY.
FT DISULFID 376 387 BY SIMILARITY.
FT DISULFID 381 398 BY SIMILARITY.
FT DISULFID 400 409 BY SIMILARITY.
FT DISULFID 416 429 BY SIMILARITY.
FT DISULFID 423 438 BY SIMILARITY.
FT DISULFID 440 449 BY SIMILARITY.
FT DISULFID 456 467 BY SIMILARITY.
FT DISULFID 461 476 BY SIMILARITY.
FT DISULFID 478 487 BY SIMILARITY.
FT DISULFID 494 505 BY SIMILARITY.
FT DISULFID 499 514 BY SIMILARITY.
FT DISULFID 516 525 BY SIMILARITY.
FT DISULFID 532 543 BY SIMILARITY.
FT DISULFID 537 552 BY SIMILARITY.
FT DISULFID 554 563 BY SIMILARITY.
FT DISULFID 570 580 BY SIMILARITY.
FT DISULFID 575 589 BY SIMILARITY.
FT DISULFID 591 600 BY SIMILARITY.
FT DISULFID 607 618 BY SIMILARITY.
FT DISULFID 612 627 BY SIMILARITY.
FT DISULFID 629 638 BY SIMILARITY.
FT DISULFID 645 655 BY SIMILARITY.
FT DISULFID 650 664 BY SIMILARITY.
FT DISULFID 666 675 BY SIMILARITY.
FT DISULFID 682 693 BY SIMILARITY.
FT DISULFID 687 702 BY SIMILARITY.
FT DISULFID 704 713 BY SIMILARITY.
FT DISULFID 720 730 BY SIMILARITY.
FT DISULFID 725 739 BY SIMILARITY.
FT DISULFID 741 750 BY SIMILARITY.
FT DISULFID 757 768 BY SIMILARITY.
FT DISULFID 762 777 BY SIMILARITY.
FT DISULFID 779 788 BY SIMILARITY.
FT DISULFID 795 806 BY SIMILARITY.

Query Match 44.7% Score 42; DB 1; Length 2444;
 Best Local Similarity 63.6% Pred. No. 1.2e+02;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 912 CRPNCHNGGS 922

4 CRPNFYNGGS 14
 1:11:1111

NTCL_MOUSE STANDARD: PRF; 2531 AA.

AC 001705:
 01-NOV-1995 (Rel. 32, Created)
 01-FEB-1996 (Rel. 33, Last sequence update)
 15-JUL-1999 (Rel. 38, Last annotation update)
 DE NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR (NOTCH PROTEIN).
 GN NOTCH1 OR NOTCH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=EMBRYO;
 RX MEDLINE=93194170; PubMed=8449489;
 RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,
 RA Copeland N.G., Gridley T.;
 RT "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
 RT homolog of Drosophila Notch.";
 RL Genomics 15:259-264(1993).
 RN [2]
 RP SEQUENCE OF 1551-2170 FROM N.A.
 RC TISSUE=EMBRYO;
 RX MEDLINE=93048835; PubMed=1425352;
 RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
 RA Greenspan R.J., McMahon A.P., Gridley T.;
 RT "Expression pattern of Notch, a mouse homolog of Drosophila Notch,
 RT suggests an important role in early postimplantation mouse
 RT development.";
 RL Development 115:737-744(1992).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
 CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL: 211886; CAA77941.1; -
 DR HSPD; P00740; 11XA.
 DR MGD; MGI:97363; NOTCH1.
 DR INTERPRO: IPR000152; -
 DR INTERPRO: IPR000561; -
 DR INTERPRO: IPR000800; -
 DR INTERPRO: IPR001438; -
 DR INTERPRO: IPR001861; -
 DR INTERPRO: IPR002110; -
 DR PFAM; PF00008; EGF; 35.
 DR PFAM; PF00023; ank; 6.
 DR PFAM; PF00066; notch; 3.
 DR PRINTS: PR00010; EGFBLD.
 DR PROSITE; PS50088; ANK_REPEAT; 2.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 27.
 DR PROSITE; PS01187; EGF_CA; 21.
 DR Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Signal; Glycoprotein.
 FT SIGNAL 1 18
 FT CHAIN 19 2531
 FT DOMAIN 19 1725
 FT TRANSMEM 1726 1746
 FT DOMAIN 1747 2531
 FT DOMAIN 20 1426
 FT DOMAIN 20 58
 FT DOMAIN 59 99
 FT DOMAIN 102 139
 FT DOMAIN 140 176
 FT DOMAIN 178 216
 FT DOMAIN 218 255
 FT DOMAIN 257 293
 FT DOMAIN 295 333
 FT DOMAIN 335 371
 FT DOMAIN 372 410
 FT DOMAIN 412 450
 FT DOMAIN 452 488
 FT DOMAIN 490 526
 FT DOMAIN 528 564
 FT DOMAIN 566 601
 FT DOMAIN 603 639
 FT DOMAIN 641 676
 FT DOMAIN 678 714
 FT DOMAIN 716 751
 FT DOMAIN 753 789
 FT DOMAIN 791 827
 FT DOMAIN 829 867
 FT DOMAIN 869 905
 FT DOMAIN 907 943
 FT DOMAIN 945 981
 FT DOMAIN 1021 1057
 FT DOMAIN 1059 1095
 FT DOMAIN 1097 1143
 FT DOMAIN 1145 1181
 FT DOMAIN 1183 1219
 FT DOMAIN 1221 1265
 FT DOMAIN 1267 1305
 FT DOMAIN 1307 1346
 FT DOMAIN 1348 1384
 FT DOMAIN 1387 1426
 FT DOMAIN 1429 1462
 FT DOMAIN 1449 1485
 FT DOMAIN 1445 1562
 FT REPEAT 1445 1480
 FT REPEAT 1481 1522
 FT REPEAT 1523 1562
 FT DOMAIN 1865 2075

NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
 EXTRACELLULAR (POTENTIAL).
 POTENTIAL.
 CYTOPLASMIC (POTENTIAL).
 36 X EGF-TYPE REPEATS.
 EGF-LIKE 1.
 EGF-LIKE 2.
 EGF-LIKE 3.
 EGF-LIKE 4.
 EGF-LIKE 5.
 EGF-LIKE 6.
 EGF-LIKE 7.
 EGF-LIKE 8.
 EGF-LIKE 9.
 EGF-LIKE 10.
 EGF-LIKE 11.
 EGF-LIKE 12.
 EGF-LIKE 13.
 EGF-LIKE 14.
 EGF-LIKE 15.
 EGF-LIKE 16.
 EGF-LIKE 17.
 EGF-LIKE 18.
 EGF-LIKE 19.
 EGF-LIKE 20.
 EGF-LIKE 21.
 EGF-LIKE 22.
 EGF-LIKE 23.
 EGF-LIKE 24.
 EGF-LIKE 25.
 EGF-LIKE 26.
 EGF-LIKE 27.
 EGF-LIKE 28.
 EGF-LIKE 29.
 EGF-LIKE 30.
 EGF-LIKE 31.
 EGF-LIKE 32.
 EGF-LIKE 33.
 EGF-LIKE 34.
 EGF-LIKE 35.
 EGF-LIKE 36.
 CYS-RICH.
 3 X LIN/NOTCH REPEATS.
 LIN/NOTCH 1.
 LIN/NOTCH 2.
 LIN/NOTCH 3.
 6 X ANK MOTIF REPEATS.

```

FT REPEAT 1865 1910 ANK MOTIF 1.
FT REPEAT 1912 1942 ANK MOTIF 2.
FT REPEAT 1944 1975 ANK MOTIF 3.
FT REPEAT 1978 2009 ANK MOTIF 4.
FT REPEAT 2011 2042 ANK MOTIF 5.
FT REPEAT 2044 2075 ANK MOTIF 6.
FT DISULFID 37 BY SIMILARITY.
FT DISULFID 31 46 BY SIMILARITY.
FT DISULFID 63 74 BY SIMILARITY.
FT DISULFID 68 87 BY SIMILARITY.
FT DISULFID 89 98 BY SIMILARITY.
FT DISULFID 106 117 BY SIMILARITY.
FT DISULFID 111 127 BY SIMILARITY.
FT DISULFID 129 138 BY SIMILARITY.
FT DISULFID 144 155 BY SIMILARITY.
FT DISULFID 149 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 182 195 BY SIMILARITY.
FT DISULFID 189 204 BY SIMILARITY.
FT DISULFID 206 215 BY SIMILARITY.
FT DISULFID 222 233 BY SIMILARITY.
FT DISULFID 227 243 BY SIMILARITY.
FT DISULFID 245 254 BY SIMILARITY.
FT DISULFID 261 272 BY SIMILARITY.
FT DISULFID 266 281 BY SIMILARITY.
FT DISULFID 283 292 BY SIMILARITY.
FT DISULFID 299 312 BY SIMILARITY.
FT DISULFID 306 321 BY SIMILARITY.
FT DISULFID 323 332 BY SIMILARITY.
FT DISULFID 339 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 361 370 BY SIMILARITY.
FT DISULFID 376 387 BY SIMILARITY.
FT DISULFID 381 398 BY SIMILARITY.
FT DISULFID 400 409 BY SIMILARITY.
FT DISULFID 416 429 BY SIMILARITY.
FT DISULFID 423 438 BY SIMILARITY.
FT DISULFID 440 449 BY SIMILARITY.
FT DISULFID 456 467 BY SIMILARITY.
FT DISULFID 461 476 BY SIMILARITY.
FT DISULFID 478 487 BY SIMILARITY.
FT DISULFID 494 505 BY SIMILARITY.
FT DISULFID 499 514 BY SIMILARITY.
FT DISULFID 516 525 BY SIMILARITY.
FT DISULFID 532 543 BY SIMILARITY.
FT DISULFID 537 552 BY SIMILARITY.
FT DISULFID 554 563 BY SIMILARITY.
FT DISULFID 570 580 BY SIMILARITY.
FT DISULFID 575 589 BY SIMILARITY.
FT DISULFID 591 600 BY SIMILARITY.
FT DISULFID 607 618 BY SIMILARITY.
FT DISULFID 612 627 BY SIMILARITY.
FT DISULFID 629 638 BY SIMILARITY.
FT DISULFID 645 655 BY SIMILARITY.
FT DISULFID 650 664 BY SIMILARITY.
FT DISULFID 666 675 BY SIMILARITY.
FT DISULFID 682 693 BY SIMILARITY.
FT DISULFID 687 702 BY SIMILARITY.
FT DISULFID 704 713 BY SIMILARITY.
FT DISULFID 720 730 BY SIMILARITY.
FT DISULFID 725 739 BY SIMILARITY.
FT DISULFID 741 750 BY SIMILARITY.
FT DISULFID 757 768 BY SIMILARITY.
FT DISULFID 762 777 BY SIMILARITY.
FT DISULFID 779 788 BY SIMILARITY.
FT DISULFID 795 806 BY SIMILARITY.
FT DISULFID 800 815 BY SIMILARITY.
FT DISULFID 817 826 BY SIMILARITY.
FT DISULFID 833 844 BY SIMILARITY.
FT DISULFID 838 855 BY SIMILARITY.
FT DISULFID 857 866 BY SIMILARITY.
FT DISULFID 873 884 BY SIMILARITY.
FT DISULFID 878 893 BY SIMILARITY.

```

```

FT DISULFID 895 904 BY SIMILARITY.
FT DISULFID 911 922 BY SIMILARITY.
FT DISULFID 916 931 BY SIMILARITY.
FT DISULFID 933 942 BY SIMILARITY.
FT DISULFID 987 998 BY SIMILARITY.
FT DISULFID 992 1007 BY SIMILARITY.
FT DISULFID 1009 1018 BY SIMILARITY.
FT DISULFID 1025 1036 BY SIMILARITY.
FT DISULFID 1030 1045 BY SIMILARITY.
FT DISULFID 1047 1056 BY SIMILARITY.
FT DISULFID 1063 1074 BY SIMILARITY.
FT DISULFID 1068 1083 BY SIMILARITY.
FT DISULFID 1085 1094 BY SIMILARITY.
FT DISULFID 1101 1122 BY SIMILARITY.

Query Match 44.7%: Score 42; DB 1; Length 2531;
Best Local Similarity 63.6%: Pred. NO. 1.3e+02;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 CKPNFYNGS 14
Db 911 CRPNCHNGS 921

RESULT 10
ID NCBI RAT STANDARD; PRT; 2531 AA.
NCBI RAT
AC 007008;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR.
GN NOTCH1.
OS Rattus norvegicus (Rat).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CC [1]
RC TISSUE-SCHMANN CELL.
RX MEDLINE=92111383; PubMed=1764995;
RA Weinmaster G., Roberts V.J., Lemke G.;
RT "A homolog of Drosophila Notch expressed during mammalian
development.";
RL development 113:199-205(1991).
RN [2]
RP REVISIONS TO 1652-1653.
RA Weinmaster G.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR THE CORRECT DIFFERENTIATION OF A NUMBER
CC OF TISSUES.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- DEVELOPMENTAL STAGE: IN THE EMBRYO, HIGHEST LEVELS OCCUR BETWEEN
CC DAYS 12 AND 14 AND DECREASE RAPIDLY TO MUCH LOWER LEVELS IN THE
CC ADULT.
CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X57405; CAA40667.1; -.
CC HSSP: P00740; ITXA.
CC INTERPRO: IPR000152; -.
CC INTERPRO: IPR000561; -.
CC INTERPRO: IPR000800; -.
CC INTERPRO: IPR001438; -.

```

[illegible]

FT	DISULEID	129	138	BY SIMILARITY
FT	DISULEID	144	155	BY SIMILARITY
FT	DISULEID	149	164	BY SIMILARITY
FT	DISULEID	166	175	BY SIMILARITY
FT	DISULEID	182	195	BY SIMILARITY
FT	DISULEID	189	204	BY SIMILARITY
FT	DISULEID	206	215	BY SIMILARITY
FT	DISULEID	222	233	BY SIMILARITY
FT	DISULEID	227	243	BY SIMILARITY
FT	DISULEID	245	254	BY SIMILARITY
FT	DISULEID	261	272	BY SIMILARITY
FT	DISULEID	266	281	BY SIMILARITY
FT	DISULEID	283	292	BY SIMILARITY
FT	DISULEID	299	312	BY SIMILARITY
FT	DISULEID	306	321	BY SIMILARITY
FT	DISULEID	323	332	BY SIMILARITY
FT	DISULEID	339	350	BY SIMILARITY
FT	DISULEID	344	359	BY SIMILARITY
FT	DISULEID	361	370	BY SIMILARITY
FT	DISULEID	376	387	BY SIMILARITY
FT	DISULEID	381	398	BY SIMILARITY
FT	DISULEID	400	409	BY SIMILARITY
FT	DISULEID	416	429	BY SIMILARITY
FT	DISULEID	423	438	BY SIMILARITY
FT	DISULEID	440	449	BY SIMILARITY
FT	DISULEID	456	467	BY SIMILARITY
FT	DISULEID	461	476	BY SIMILARITY
FT	DISULEID	478	487	BY SIMILARITY
FT	DISULEID	494	505	BY SIMILARITY
FT	DISULEID	499	514	BY SIMILARITY
FT	DISULEID	516	525	BY SIMILARITY
FT	DISULEID	532	543	BY SIMILARITY
FT	DISULEID	537	552	BY SIMILARITY
FT	DISULEID	554	563	BY SIMILARITY
FT	DISULEID	570	580	BY SIMILARITY
FT	DISULEID	575	589	BY SIMILARITY
FT	DISULEID	591	600	BY SIMILARITY
FT	DISULEID	607	618	BY SIMILARITY
FT	DISULEID	612	627	BY SIMILARITY
FT	DISULEID	629	638	BY SIMILARITY
FT	DISULEID	645	655	BY SIMILARITY
FT	DISULEID	650	664	BY SIMILARITY
FT	DISULEID	666	675	BY SIMILARITY
FT	DISULEID	682	693	BY SIMILARITY
FT	DISULEID	687	702	BY SIMILARITY
FT	DISULEID	704	713	BY SIMILARITY
FT	DISULEID	720	730	BY SIMILARITY
FT	DISULEID	725	739	BY SIMILARITY
FT	DISULEID	741	750	BY SIMILARITY
FT	DISULEID	757	768	BY SIMILARITY
FT	DISULEID	762	777	BY SIMILARITY
FT	DISULEID	779	788	BY SIMILARITY
FT	DISULEID	795	806	BY SIMILARITY
FT	DISULEID	800	815	BY SIMILARITY
FT	DISULEID	817	826	BY SIMILARITY
FT	DISULEID	833	844	BY SIMILARITY
FT	DISULEID	838	855	BY SIMILARITY
FT	DISULEID	857	866	BY SIMILARITY
FT	DISULEID	873	884	BY SIMILARITY
FT	DISULEID	878	893	BY SIMILARITY
FT	DISULEID	895	904	BY SIMILARITY
FT	DISULEID	911	922	BY SIMILARITY
FT	DISULEID	916	931	BY SIMILARITY
FT	DISULEID	933	942	BY SIMILARITY
FT	DISULEID	987	998	BY SIMILARITY
FT	DISULEID	992	1007	BY SIMILARITY
FT	DISULEID	1009	1018	BY SIMILARITY
FT	DISULEID	1025	1036	BY SIMILARITY
FT	DISULEID	1030	1045	BY SIMILARITY
FT	DISULEID	1047	1056	BY SIMILARITY
FT	DISULEID	1064	1074	BY SIMILARITY
FT	DISULEID	1068	1083	BY SIMILARITY
FT	DISULEID	1085	1094	BY SIMILARITY

Query Match 44.7%; Score 42; DB 1; Length 2531;
 Best Local Similarity 63.6%; Pred. No. 1,3e+02;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 4 CRPNFYNGGS 14
 Db 911 CRPNCHNGGS 921

RESULT 11
 LMA1_HUMAN STANDARD; PRT; 3075 AA.

AC P25391;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE LAMININ ALPHA-1 CHAIN PRECURSOR (LAMININ A CHAIN).
 GN LAMA1 OR LAMA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-9133420; PubMed-1714537;
 RA Haaparaanta T., Uitto J., Ruoslahti E., Engvall E.;
 RT "Molecular cloning of the cdna encoding human laminin A chain.";
 RL Matrix 11:151-160(1991).
 [2]
 RP SEQUENCE OF 1-2628 FROM N.A.
 RX MEDLINE-91264789; PubMed-2049067;
 RA Nissinen M., Vuolteenaho R., Boot-Handford R., Kallunki P.,
 RA Tryggvason K.;
 RT "Primary structure of the human laminin A chain. Limited expression
 RT in human tissues.";
 RL Biochem. J. 276:369-379(1991).
 RN [3]
 RP SEQUENCE OF 2397-3072 FROM N.A.
 RX MEDLINE-89280632; PubMed-273383;
 RA Olsen D., Nagayoshi T., Fazio M., Peltonen J., Jaakkola S.,
 RA Sanborn D., Sasaki T., Kuitavien H., Chu M.L., Deutzmann R.,
 RA Timpl R., Uitto J.;
 RT "Human laminin: cloning and sequence analysis of cdnas encoding A, B1
 RT and B2 chains, and expression of the corresponding genes in human
 RT skin and cultured cells.";
 RL Lab. Invest. 60:772-782(1989).
 CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR. LAMININ
 CC IS THOUGHT TO MEDATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
 CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
 CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
 CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
 CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
 CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
 CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOULES AT EACH END.
 CC THE ALPHA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ) AND
 CC LAMININ-3 (S-LAMININ).
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
 CC COMPONENT).
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -1- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
 CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
 CC -1- SIMILARITY: CONTAINS 17 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
 CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X58531; CAA1418.1; -
 DR PIR: S14458; S14458.
 DR HSP: P02468; TITLE.
 DR MIM: 150320; -
 DR INTERPRO: IPR000034; -
 DR INTERPRO: IPR000561; -
 DR INTERPRO: IPR001791; -
 DR INTERPRO: IPR001886; -
 DR INTERPRO: IPR002049; -
 DR PFAM: PF00052; laminin_B; 2.
 DR PFAM: PF00053; laminin_EGF; 15.
 DR PFAM: PF00054; laminin_G; 5.
 DR PFAM: PF00055; laminin_Nterm; 1.
 DR PRINTS: PR00011; EGF_LAMININ.
 DR PROSITE: PS00022; EGF_1; 11.
 DR PROSITE: PS01186; EGF_2; 2.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; 15.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; 15.
 DR Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 DR laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 FT SIGNAL 1 17
 FT CHAIN 18 3075
 FT DOMAIN 18 269
 FT DOMAIN 270 516
 FT
 FT DOMAIN 270 326
 FT DOMAIN 327 396
 FT DOMAIN 397 453
 FT DOMAIN 454 502
 FT DOMAIN 503 512
 FT DOMAIN 517 708
 FT DOMAIN 709 1159
 FT
 FT DOMAIN 709 741
 FT DOMAIN 742 790
 FT DOMAIN 791 848
 FT DOMAIN 849 901
 FT DOMAIN 902 950
 FT DOMAIN 951 997
 FT DOMAIN 998 1043
 FT DOMAIN 1044 1089
 FT DOMAIN 1090 1149
 FT DOMAIN 1150 1159
 FT DOMAIN 1160 1361
 FT DOMAIN 1362 1555
 FT
 FT DOMAIN 1362 1402
 FT DOMAIN 1403 1451
 FT DOMAIN 1452 1508
 FT DOMAIN 1509 1555
 FT DOMAIN 1556 2025
 FT DOMAIN 2026 2125
 FT DOMAIN 2126 2327
 FT DOMAIN 2328 2509
 FT DOMAIN 2510 2736
 FT DOMAIN 2737 2913
 FT DOMAIN 2914 3075
 FT DOMAIN 1706 1796
 FT DOMAIN 1968 1989
 FT DOMAIN 2088 2120
 FT DOMAIN 2534 2536
 FT SITE 270 279
 FT DISULFID 292 290
 FT DISULFID 292 301
 FT DISULFID 297 305

 LAMININ EGF-LIKE 5 (C-TERMINAL).
 LAMININ EGF-LIKE 6.
 LAMININ EGF-LIKE 7.
 LAMININ EGF-LIKE 8.
 LAMININ EGF-LIKE 9.
 LAMININ EGF-LIKE 10.
 LAMININ EGF-LIKE 11.
 LAMININ EGF-LIKE 12.
 LAMININ EGF-LIKE 13.
 LAMININ EGF-LIKE 14 (N-TERMINAL).
 LAMININ DOMAIN IV 2 (DOMAIN IV A).
 9 X LAMININ EGF-LIKE REPEATS (DOMAIN
 III B).
 LAMININ EGF-LIKE 5 (C-TERMINAL).
 LAMININ EGF-LIKE 6.
 LAMININ EGF-LIKE 7.
 LAMININ EGF-LIKE 8.
 LAMININ EGF-LIKE 9.
 LAMININ EGF-LIKE 10.
 LAMININ EGF-LIKE 11.
 LAMININ EGF-LIKE 12.
 LAMININ EGF-LIKE 13.
 LAMININ EGF-LIKE 14 (N-TERMINAL).
 LAMININ DOMAIN IV 2 (DOMAIN IV A).
 3.5 X LAMININ EGF-LIKE REPEATS (DOMAIN
 III A).
 LAMININ EGF-LIKE 14 (C-TERMINAL).
 LAMININ EGF-LIKE 15.
 LAMININ EGF-LIKE 16.
 LAMININ EGF-LIKE 17.
 DOMAIN II AND I.
 5 X LAMININ G-LIKE REPEATS (DOMAIN G).
 LAMININ G-LIKE 1.
 LAMININ G-LIKE 2.
 LAMININ G-LIKE 3.
 LAMININ G-LIKE 4.
 LAMININ G-LIKE 5.
 COILED COIL (POTENTIAL).
 COILED COIL (POTENTIAL).
 CELL ATTACHMENT SITE.
 BY SIMILARITY.
 BY SIMILARITY.
 POTENTIAL.

```

FT DISULFID 304 324 BY SIMILARITY.
FT DISULFID 327 336 BY SIMILARITY.
FT DISULFID 329 361 BY SIMILARITY.
FT DISULFID 364 373 BY SIMILARITY.
FT DISULFID 376 394 BY SIMILARITY.
FT DISULFID 397 409 BY SIMILARITY.
FT DISULFID 427 438 BY SIMILARITY.
FT DISULFID 429 438 BY SIMILARITY.
FT DISULFID 441 451 BY SIMILARITY.
FT DISULFID 454 467 BY SIMILARITY.
FT DISULFID 456 471 BY SIMILARITY.
FT DISULFID 473 482 BY SIMILARITY.
FT DISULFID 485 500 BY SIMILARITY.
FT DISULFID 742 751 BY SIMILARITY.
FT DISULFID 744 757 BY SIMILARITY.
FT DISULFID 760 769 BY SIMILARITY.
FT DISULFID 772 788 BY SIMILARITY.
FT DISULFID 791 806 BY SIMILARITY.
FT DISULFID 793 816 BY SIMILARITY.
FT DISULFID 819 828 BY SIMILARITY.
FT DISULFID 831 846 BY SIMILARITY.
FT DISULFID 849 863 BY SIMILARITY.
FT DISULFID 851 870 BY SIMILARITY.
FT DISULFID 873 882 BY SIMILARITY.
FT DISULFID 885 899 BY SIMILARITY.
FT DISULFID 902 914 BY SIMILARITY.
FT DISULFID 904 921 BY SIMILARITY.
FT DISULFID 923 932 BY SIMILARITY.
FT DISULFID 935 948 BY SIMILARITY.
FT DISULFID 951 963 BY SIMILARITY.
FT DISULFID 953 969 BY SIMILARITY.
FT DISULFID 971 980 BY SIMILARITY.
FT DISULFID 983 995 BY SIMILARITY.
FT DISULFID 998 1007 BY SIMILARITY.
FT DISULFID 1000 1014 BY SIMILARITY.
FT DISULFID 1016 1025 BY SIMILARITY.
FT DISULFID 1028 1041 BY SIMILARITY.
FT DISULFID 1044 1056 BY SIMILARITY.
FT DISULFID 1063 1074 BY SIMILARITY.
FT DISULFID 1077 1087 BY SIMILARITY.
FT DISULFID 1403 1412 BY SIMILARITY.
FT DISULFID 1405 1419 BY SIMILARITY.
FT DISULFID 1422 1431 BY SIMILARITY.
FT DISULFID 1434 1449 BY SIMILARITY.
FT DISULFID 1452 1466 BY SIMILARITY.
FT DISULFID 1454 1476 BY SIMILARITY.
FT DISULFID 1479 1488 BY SIMILARITY.
FT DISULFID 1491 1506 BY SIMILARITY.
FT DISULFID 1509 1521 BY SIMILARITY.
FT DISULFID 1511 1528 BY SIMILARITY.
FT DISULFID 1530 1539 BY SIMILARITY.
FT DISULFID 1542 1553 BY SIMILARITY.
FT DISULFID 1556 1556 INTERCHAIN (PROBABLE).
FT DISULFID 1560 1560 INTERCHAIN (PROBABLE).
FT CARBOHYD 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 555 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 665 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 763 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 926 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 952 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1045 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1407 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1479 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1579 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1596 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1678 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1689 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1698 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1717 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1804 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1894 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1898 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1957 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

Query Match 44.7%; Score 42; DB 1; Length 3075;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CYCKRPNFY 9
DB 482 CDRCKRPFY 490

RESULT 12
LMAI_MOUSE STANDARD; PRI: 3084 AA.
ID LMAI_MOUSE P19137;
AC 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE LAMININ ALPHA-1 CHAIN PRECURSOR (LAMININ A CHAIN).
GN LAMA1 OR LAMA-1 OR LAMA.
OS Mus musculus (Mouse).
OC Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
RN [1]
RP MEDLINE=89034134; PubMed=3182802;
RX Sasaki M., Kleinman H.K., Huber H., Deutzmann R., Yamada Y.;
RT "Laminin, a multidomain protein. The A chain has a unique globular
domain and homology with the basement membrane proteoglycan and the
laminin B chains."
RL J. Biol. Chem. 263:16536-16544(1988).
RN [2]
RP SEQUENCE OF 1-339 FROM N.A.
RX MEDLINE=88225080; PubMed=3267223;
RA Hartl L., Oberbauer I., Deutzmann R.;
RT "The N terminus of laminin A chain is homologous to the B chains."
RL Eur. J. Biochem. 173:629-635(1988).
RN [3]
RP SEQUENCE OF 2538-3084 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=89030693; PubMed=3181157;
RA Deutzmann R., Huber J., Schmetz K.A., Oberbauer I., Hartl L.;
RT "Structural study of long arm fragments of laminin. Evidence for
repetitive C-terminal sequences in the A-chain, not present in the B-
chains."
RL Eur. J. Biochem. 177:35-45(1988).
CC -I- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR. LAMININ
IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
CC -I- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
CC THE ALPHA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ) AND
LAMININ-3 (S-LAMININ).
CC LAMININ-3 (S-LAMININ).
CC -I- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -I- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
COMPONENT).
CC -I- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -I- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
CC -I- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -I- SIMILARITY: CONTAINS 17 LAMININ EGF-LIKE DOMAINS.
CC -I- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
CC -I- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announcel/
or send an email to license@isb-sib.ch).
CC

```

DR	EMBL: J04064	AAA39410.1	-	
DR	EMBL: X07737	CAA30561.1	-	
DR	EMBL: X13459	CAA31807.1	-	
DR	EMBL: M36775	AAA39406.1	-	
DR	PIR: A31771	MMMSA-		
DR	HSSP: P02468	1TLE		
DR	MGI: 99892	LAM1		
DR	INTERPRO: IPR000034	-		
DR	INTERPRO: IPR000561	-		
DR	INTERPRO: IPR001791	-		
DR	INTERPRO: IPR001866	-		
DR	INTERPRO: IPR002049	-		
DR	PFAM: PF00052	Laminin_B_2		
DR	PFAM: PF00053	Laminin_BGF_15		
DR	PFAM: PF00054	Laminin_BGF_5		
DR	PFAM: PF00055	Laminin_Nterm_1		
DR	PRINTS: PR00011	EGFLAMININ		
DR	PROSITE: PS00022	EGF_1: 11		
DR	PROSITE: PS0186	EGF_2: 3		
DR	PROSITE: PS01248	LAMININ_TYPE_EGF_15		
KW	Glycoprotein	Basement membrane; Cell adhesion; Repat; Signal.		
KW	Laminin EGF-like domain	POTENTIAL		
FT	CHAIN	1 24		
FT	MOD_RES	25 3084		
FT	DOMAIN	25 276		
FT	DOMAIN	277 519		
FT	DOMAIN	277 333		
FT	DOMAIN	334 403		
FT	DOMAIN	404 460		
FT	DOMAIN	461 509		
FT	DOMAIN	510 519		
FT	DOMAIN	520 715		
FT	DOMAIN	716 1166		
FT	DOMAIN	716 748		
FT	DOMAIN	749 797		
FT	DOMAIN	798 855		
FT	DOMAIN	856 908		
FT	DOMAIN	909 957		
FT	DOMAIN	958 1004		
FT	DOMAIN	1005 1050		
FT	DOMAIN	1051 1096		
FT	DOMAIN	1097 1156		
FT	DOMAIN	1157 1166		
FT	DOMAIN	1167 1368		
FT	DOMAIN	1369 1561		
FT	DOMAIN	1369 1409		
FT	DOMAIN	1410 1458		
FT	DOMAIN	1459 1515		
FT	DOMAIN	1516 1562		
FT	DOMAIN	1562 2133		
FT	DOMAIN	2134 3084		
FT	DOMAIN	2148 2335		
FT	DOMAIN	2336 2517		
FT	DOMAIN	2518 2745		
FT	DOMAIN	2746 2922		
FT	DOMAIN	2923 3084		
FT	DOMAIN	1612 1820		
FT	DOMAIN	1869 1903		
FT	DOMAIN	2096 2128		
FT	SITE	1147 1149		
FT	DISULFID	277 286		
FT	DISULFID	279 297		
FT	DISULFID	299 308		
FT	DISULFID	311 331		
FT	DISULFID	334 343		
FT	DISULFID	336 368		
FT	DISULFID	371 380		
FT	DISULFID	383 401		
FT	DISULFID	404 416		
FT	DISULFID	404 416		

FT	DISULFID	406	434	BY SIMILARITY.
FT	DISULFID	436	445	BY SIMILARITY.
FT	DISULFID	448	458	BY SIMILARITY.
FT	DISULFID	461	474	BY SIMILARITY.
FT	DISULFID	463	478	BY SIMILARITY.
FT	DISULFID	480	489	BY SIMILARITY.
FT	DISULFID	492	507	BY SIMILARITY.
FT	DISULFID	749	758	BY SIMILARITY.
FT	DISULFID	751	764	BY SIMILARITY.
FT	DISULFID	767	776	BY SIMILARITY.
FT	DISULFID	779	795	BY SIMILARITY.
FT	DISULFID	798	813	BY SIMILARITY.
FT	DISULFID	800	823	BY SIMILARITY.
FT	DISULFID	826	835	BY SIMILARITY.
FT	DISULFID	838	853	BY SIMILARITY.
FT	DISULFID	856	870	BY SIMILARITY.
FT	DISULFID	858	877	BY SIMILARITY.
FT	DISULFID	880	889	BY SIMILARITY.
FT	DISULFID	892	906	BY SIMILARITY.
FT	DISULFID	909	921	BY SIMILARITY.
FT	DISULFID	911	928	BY SIMILARITY.
FT	DISULFID	930	939	BY SIMILARITY.
FT	DISULFID	942	955	BY SIMILARITY.
FT	DISULFID	958	970	BY SIMILARITY.
FT	DISULFID	960	976	BY SIMILARITY.
FT	DISULFID	978	987	BY SIMILARITY.
FT	DISULFID	990	1002	BY SIMILARITY.
FT	DISULFID	1005	1014	BY SIMILARITY.
FT	DISULFID	1007	1021	BY SIMILARITY.
FT	DISULFID	1023	1032	BY SIMILARITY.
FT	DISULFID	1035	1048	BY SIMILARITY.
FT	DISULFID	1051	1063	BY SIMILARITY.
FT	DISULFID	1072	1081	BY SIMILARITY.
FT	DISULFID	1084	1094	BY SIMILARITY.
FT	DISULFID	1410	1419	BY SIMILARITY.
FT	DISULFID	1412	1426	BY SIMILARITY.
FT	DISULFID	1429	1438	BY SIMILARITY.
FT	DISULFID	1441	1456	BY SIMILARITY.
FT	DISULFID	1459	1473	BY SIMILARITY.
FT	DISULFID	1461	1483	BY SIMILARITY.
FT	DISULFID	1486	1495	BY SIMILARITY.
FT	DISULFID	1498	1513	BY SIMILARITY.
FT	DISULFID	1516	1528	BY SIMILARITY.
FT	DISULFID	1518	1535	BY SIMILARITY.
FT	DISULFID	1537	1546	BY SIMILARITY.
FT	DISULFID	1549	1560	BY SIMILARITY.
FT	DISULFID	1563	1563	BY SIMILARITY.
FT	DISULFID	1567	1567	INTERCHAIN (PROBABLE).
FT	CAROHND	45	45	INTERCHAIN (PROBABLE).
FT	CAROHND	79	79	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CAROHND	370	370	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CAROHND	374	374	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CAROHND	531	531	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CAROHND	562	562	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CAROHND	672	672	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CAROHND	770	770	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CAROHND	857	857	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CAROHND	914	914	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CAROHND	959	959	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CAROHND	969	969	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CAROHND	1052	1052	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CAROHND	1344	1344	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CAROHND	1414	1414	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CAROHND	1566	1566	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CAROHND	1603	1603	N-LINKED (GLCNAC. . .) (POTENTIAL).
Query Match				
Best Local Similarity 66.7%;				
Matches 677 Conservative 1; Mismatches 2; Indels 0; Gaps 0.				
Score 42; DB 1; Length 3084;				
44.7%;				
1 CVCKRPNFY 9				
I : : : : : I I I				

Db 489 CDRCKPGEF 497

RESULT 13

LMA_DROME STANDARD: PRT: 3712 AA.

AC 000174.

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE LAMININ ALPHA CHAIN PRECURSOR.

CN LANA OR LAMA.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

NC [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93049203; PubMed=1425586;

RA Kusche-Gullberg M., Garrison K., Mackrell A.J., Fessler L.I.,

RA Fessler J.H.;

RT "Laminin A chain: expression during Drosophila development and genomic sequence."

RL EMBO J. 11:4519-4527(1992).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=EMBRYO;

RX MEDLINE=94038678; PubMed=8223265;

RA Hencliffe C., Garcia-Alonso L., Tang J., Goodman C.S.;

RT "Genetic analysis of laminin A reveals diverse functions during morphogenesis in Drosophila."

RL Development 118:325-337(1993).

RN [3]

RP SEQUENCE OF 1762-3712 FROM N.A.

RX MEDLINE=92078147; PubMed=1744083;

RA Garrison K., Mackrell A.J., Fessler J.H.;

RT "Drosophila laminin A chain sequence, interspecies comparison, and RT domain structure of a major carboxyl portion."

RL J. Biol. Chem. 266:22899-22904(1991).

CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR. LAMININ IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.

CC -1- FUNCTION: DIVERSE FUNCTIONS DURING MORPHOGENESIS IN DROSOPHILA. COMPLETE LOSS-OF-FUNCTION MUTATIONS LEAD TO LATE EMBRYONIC LETHALITY. CERTAIN PARTIAL LOSS-OF-FUNCTION MUTATIONS GIVE RAISE TO ESCAPER ADULTS, WHICH HAVE ROUGH EYES ASSOCIATED WITH CHANGES IN CELL FATE AND PATTERN, MISSHAPEN LEGS AND DEFECTS IN WING STRUCTURE.

CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.

CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT MEMBRANES (MAJOR COMPONENT).

CC -1- TISSUE SPECIFICITY: NEWLY FORMED MESODERM AND LATER PROMINENTLY EXPRESSED IN HEMOCYTES, WHICH ALSO SYNTHESIZE COLLAGEN IV.

CC -1- DEVELOPMENTAL STAGE: DURING MORPHOGENESIS, MOSTLY IN EMBRYO DEVELOPMENT AT 10-12 HOURS.

CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT WITH OTHER LAMININ CHAINS TO FORM A COILED-COIL STRUCTURE.

CC -1- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.

CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).

CC -1- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.

CC -1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV (DOMAIN IV' IS NOT SIMILAR TO LAMININ DOMAIN IV).

CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: M96388; AAA28662.1; -

DR EMBL: L07288; AAC37178.1; -

DR EMBL: M75882; AAA28661.1; -

DR HSSP: P02468; TITLE.

DR FLYBASE: FBgn0002526; LANA.

DR INTERPRO: IPR000034; -

DR INTERPRO: IPR000561; -

DR INTERPRO: IPR001791; -

DR INTERPRO: IPR001866; -

DR INTERPRO: IPR002049; -

DR PFAM: PF000052; laminin_B; 1.

DR PFAM: PF000053; laminin_EGF; 20.

DR PFAM: PF000054; laminin_G; 5.

DR PFAM: PF000055; laminin_Nterm; 1.

DR PRINTS: PR00011; EGF/LAMININ.

DR PROSITE: PS00022; EGF_1; 17.

DR PROSITE: PS01186; EGF_2; 5.

DR PROSITE: PS01248; LAMININ TYPE EGF; 19.

DR Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil; Laminin EGF-like domain; Cell adhesion; Repeat; Signal.

KM Laminin EGF-like domain; Cell adhesion; Repeat; Signal.

FT SIGNAL 1 22

FT CHAIN 23 3712

FT DOMAIN 25 272

FT DOMAIN 273 815

FT 273 332

FT 333 402

FT 403 447

FT 448 494

FT 495 540

FT 541 586

FT 587 631

FT 632 676

FT 677 731

FT 732 784

FT 785 815

FT 816 1374

FT 1375 1574

FT 1375 1420

FT 1421 1465

FT 1466 1513

FT 1514 1564

FT 1565 1574

FT 1575 1775

FT 1776 2111

FT 1776 1808

FT 1809 1858

FT 1859 1916

FT 1917 1969

FT 1970 2016

FT 2017 2063

FT 2064 2111

FT 2112 2697

FT 2698 3712

FT 3712 2698

FT 2698 2862

FT 2863 3048

FT 3049 3223

FT 3223 3370

FT 3370 3296

FT 3296 3334

FT 3334 3528

FT 3529 3712

FT 3712 2249

FT 2249 2301

FT 2301 2321

FT 2321 2450

FT 2450 2541

FT 2541 2676

FT 2676 273

FT 273 282

FT 282 296

FT 296 307

FT 307 330

FT DISULFID 310

FT DISULFID 330

111 A).

4.5 X LAMININ EGF-LIKE REPEATS (DOMAIN III B).

LAMININ EGF-LIKE 12.

LAMININ EGF-LIKE 13.

LAMININ EGF-LIKE 14.

LAMININ EGF-LIKE 15.

LAMININ EGF-LIKE 16 (N-TERMINAL).

LAMININ DOMAIN IV (DOMAIN IV).

6.5 X LAMININ EGF-LIKE REPEATS (DOMAIN III A).

LAMININ EGF-LIKE 16 (C-TERMINAL).

LAMININ EGF-LIKE 17.

LAMININ EGF-LIKE 18.

LAMININ EGF-LIKE 19.

LAMININ EGF-LIKE 20.

LAMININ EGF-LIKE 21.

LAMININ EGF-LIKE 22.

LAMININ EGF-LIKE 22.

DOMAIN II AND I.

5 X LAMININ G-LIKE REPEATS (DOMAIN G).

LAMININ G-LIKE 1.

LAMININ G-LIKE 2.

LAMININ G-LIKE 3.

POLY-THR.

LAMININ G-LIKE 4.

LAMININ G-LIKE 5.

COILED COIL (POTENTIAL).

COILED COIL (POTENTIAL).

COILED COIL (POTENTIAL).

COILED COIL (POTENTIAL).

COILED COIL (POTENTIAL).

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

FT	DISULFID	333	342	BY SIMILARITY.
FT	DISULFID	335	367	BY SIMILARITY.
FT	DISULFID	370	379	BY SIMILARITY.
FT	DISULFID	382	400	BY SIMILARITY.
FT	DISULFID	403	414	BY SIMILARITY.
FT	DISULFID	405	421	BY SIMILARITY.
FT	DISULFID	423	432	BY SIMILARITY.
FT	DISULFID	435	445	BY SIMILARITY.
FT	DISULFID	448	460	BY SIMILARITY.
FT	DISULFID	450	468	BY SIMILARITY.
FT	DISULFID	470	479	BY SIMILARITY.
FT	DISULFID	482	492	BY SIMILARITY.
FT	DISULFID	495	507	BY SIMILARITY.
FT	DISULFID	497	514	BY SIMILARITY.
FT	DISULFID	516	525	BY SIMILARITY.
FT	DISULFID	528	538	BY SIMILARITY.
FT	DISULFID	541	553	BY SIMILARITY.
FT	DISULFID	543	560	BY SIMILARITY.
FT	DISULFID	562	571	BY SIMILARITY.
FT	DISULFID	574	584	BY SIMILARITY.
FT	DISULFID	587	599	BY SIMILARITY.
FT	DISULFID	589	605	BY SIMILARITY.
FT	DISULFID	607	616	BY SIMILARITY.
FT	DISULFID	619	629	BY SIMILARITY.
FT	DISULFID	632	644	BY SIMILARITY.
FT	DISULFID	634	650	BY SIMILARITY.
FT	DISULFID	652	661	BY SIMILARITY.
FT	DISULFID	664	674	BY SIMILARITY.
FT	DISULFID	677	691	BY SIMILARITY.
FT	DISULFID	679	700	BY SIMILARITY.
FT	DISULFID	702	711	BY SIMILARITY.
FT	DISULFID	714	729	BY SIMILARITY.
FT	DISULFID	732	746	BY SIMILARITY.
FT	DISULFID	734	753	BY SIMILARITY.
FT	DISULFID	755	764	BY SIMILARITY.
FT	DISULFID	767	782	BY SIMILARITY.
FT	DISULFID	1375	1387	BY SIMILARITY.
FT	DISULFID	1377	1394	BY SIMILARITY.
FT	DISULFID	1396	1405	BY SIMILARITY.
FT	DISULFID	1408	1418	BY SIMILARITY.
FT	DISULFID	1421	1429	BY SIMILARITY.
FT	DISULFID	1423	1436	BY SIMILARITY.
FT	DISULFID	1438	1447	BY SIMILARITY.
FT	DISULFID	1450	1463	BY SIMILARITY.
FT	DISULFID	1466	1480	BY SIMILARITY.
FT	DISULFID	1468	1487	BY SIMILARITY.
FT	DISULFID	1489	1498	BY SIMILARITY.
FT	DISULFID	1501	1511	BY SIMILARITY.
FT	DISULFID	1514	1526	BY SIMILARITY.
FT	DISULFID	1516	1533	BY SIMILARITY.
FT	DISULFID	1535	1544	BY SIMILARITY.
FT	DISULFID	1547	1562	BY SIMILARITY.
FT	DISULFID	1859	1874	BY SIMILARITY.
FT	DISULFID	1861	1885	BY SIMILARITY.
FT	DISULFID	1887	1896	BY SIMILARITY.
FT	DISULFID	1899	1914	BY SIMILARITY.
FT	DISULFID	1917	1931	BY SIMILARITY.
FT	DISULFID	1919	1938	BY SIMILARITY.
FT	DISULFID	1941	1950	BY SIMILARITY.

Query Match

Best Local Similarity 44.7%;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0.

OY

1 CVCKPNEY 9

111111

Db

379 CNCKPKTY 387

RESULT 14

YC81_CAEEL

ID YC81_CAEEL

QC Q19981;

STANDARD;

PRT: 1291 AA.

DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	01-NOV-1997 (Rel. 35, Last annotation update)
DE	HYPOTHETICAL 143.1 KDA PROTEIN F33C8.1 IN CHROMOSOME X PRECURSOR.
GN	F33C8.1.
OS	Caenorhabditis elegans.
OC	Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabdilitida; Rhabditoidea;
OC	Rhabdilitidae; Peloderinae; Caenorhabditis.
RP	[1]
RA	SEQUENCE FROM N.A.
RC	STRAIN-BRISTOL N2;
CC	Percy C.;
RL	Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
CC	-1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC	-1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC	-1- SIMILARITY: CONTAINS 2 LAMININ EGF-LIKE DOMAINS.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; Z69790; CA93653.1; -.
DR	HSSP; P02468; IKLO.
DR	WORMPEP; F33C8.1; CE05796.
DR	INTERPRO; IPR000561; -.
DR	INTERPRO; IPR000859; -.
DR	INTERPRO; IPR001798; -.
DR	INTERPRO; IPR002049; -.
DR	INTERPRO; IPR002165; -.
DR	PFAM; PF00431; CUB; 1.
DR	PFAM; PF01344; Kelch; 2.
DR	PFAM; PF01437; Plexin.repeat; 4.
DR	PFAM; PF00053; laminin_EGF_2.
DR	PROSITE; PS00022; EGF_L1; 2.
DR	PROSITE; PS01186; EGF_2; 2.
DR	PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR	PROSITE; PS01180; CUB; 1.
KW	Repeating protein; laminin EGF-like domain; EGF-like domain;
KM	Repeat: Signal.
FT	SIGNAL 1 ?
FT	CHAIN 1 1291 ?
FT	DOMAIN 2 92 HYPOTHETICAL 143.1 KDA PROTEIN.
FT	DOMAIN 94 200 CUB.
FT	DOMAIN 201 232 EGF-LIKE 2.
FT	DOMAIN 232 270 EGF-LIKE 3.
FT	DOMAIN 965 1019 LAMININ EGF-LIKE 1.
FT	DOMAIN 1020 1067 LAMININ EGF-LIKE 2.
FT	DISULFID 66 75 BY SIMILARITY.
FT	DISULFID 70 80 BY SIMILARITY.
FT	DISULFID 82 91 BY SIMILARITY.
FT	DISULFID 205 215 BY SIMILARITY.
FT	DISULFID 209 220 BY SIMILARITY.
FT	DISULFID 222 231 BY SIMILARITY.
FT	DISULFID 236 252 BY SIMILARITY.
FT	DISULFID 247 257 BY SIMILARITY.
FT	DISULFID 259 269 BY SIMILARITY.
FT	DISULFID 965 973 BY SIMILARITY.
FT	DISULFID 967 988 BY SIMILARITY.
FT	DISULFID 991 1000 BY SIMILARITY.
FT	DISULFID 1003 1017 BY SIMILARITY.
FT	DISULFID 1020 1029 BY SIMILARITY.
FT	DISULFID 1022 1036 BY SIMILARITY.
FT	DISULFID 1038 1048 BY SIMILARITY.
FT	DISULFID 1051 1065 BY SIMILARITY.
FT	CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 395 395 N-LINKED (GLCNAC. . .) (POTENTIAL).

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:52:54 ; Search time 116.78 Seconds
(without alignments)
15.055 Million cell updates/sec

Title: US-09-196-161D-6
Perfect score: 94
Sequence: 1 CVKCKPNEYNGGSP 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPTRMBL_15:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mhc:*
8: sp_mammal:*
9: sp_organelle:*
10: sp_phase:*
11: sp_plant:*
12: sp_rodent:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	94	100.0	394	5	Q27208
2	94	100.0	442	5	Q27208
3	49	52.1	600	5	Q9XZG2
4	49	52.1	600	5	Q9NL27
5	48	51.1	650	5	Q9NL28
6	48	51.1	401	5	Q9NH87
7	47	50.0	267	11	P97939
8	47	50.0	337	11	P97263
9	47	50.0	339	11	P97261
10	47	47.9	379	11	P70105
11	45	47.9	335	5	Q9XVB2
12	45	47.9	340	5	Q9YIA5
13	45	47.9	967	5	Q9YIA6
14	45	47.9	1571	5	Q76925
15	44	46.8	342	12	Q9YMW1
16	44	46.8	977	13	Q9YMW1
17	44	46.8	2102	5	O18183
18	44	46.8	2424	5	Q9W206
19	43	45.7	131	10	Q40415

ID	Q27208	PRELIMINARY:	PRT:	394 AA.	
AC	Q27208	01-NOV-1996 (TREMBLrel: 01, Created)			P72804 synchocyst
DT	01-NOV-1996 (TREMBLrel: 08, Last sequence update)				Q28948 archaeoglob
DT	01-JUN-2000 (TREMBLrel: 14, Last annotation update)				Q9P133 homo sapien
DE	IMMOBILIZATION ANTIGEN PRECURSOR (FRAGMENT).				Q9P144 homo sapien
OS	Ichthyophthirius multifiliis.				Q9J133 mus musculu
OC	Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;				Q9UPU3 homo sapien
OC	Opisthokonta; Ichthyophthirius.				Q9Y9K2 aeropyrum p
OX	NCBI_TaxID=5932;				Q67659 garilic late
RN	[1]				Q70712 shallot lat
RP	SEQUENCE FROM N.A.				Q70718 shallot lat
RC	STRAIN=GEORGIA;				Q70720 shallot lat
RX	MEDLINE=92335298; PubMed=1631132;				Q25664 plasmodium
RA	Clark T.G., McGraw R.A., Dickerson H.W.;				Q29208 heliobacte
RT	"Developmental expression of surface antigen genes in the parasitic				Q14293 homo sapien
RT	ciliate Ichthyophthirius multifiliis."				Q9T84 macaca fasc
RL	Proc. Natl. Acad. Sci. U.S.A. 89:6363-6367(1992).				Q9RYJ2 deinococcus
RL	[13]				Q9RW0 drosophila
RP	SEQUENCE FROM N.A.				Q9Y109 uscllago ma
RC	STRAIN=GEORGIA;				Q64756 arabidopsis
RA	Clark T.;				Q24452 drosophila
RL	Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.				Q9WZ29 drosophila
RN	[14]				Q9J180 mus musculu
RP	SEQUENCE FROM N.A.				Q9N36 homo sapien
RC	STRAIN=GEORGIA;				Q9S27 homo sapien
RA	Clark T.;				Q75445 homo sapien
RL	Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.				
DR	EMBL, M9207; AAC36158.1; .				
KW	Signal.				
FT	NON_TER				
FT	SIGNAL				

ALIGNMENTS

RESULT 1
ID Q27208 PRELIMINARY: PRT: 394 AA.
AC Q27208: 01-NOV-1996 (TREMBLrel: 01, Created)
DT 01-NOV-1996 (TREMBLrel: 08, Last sequence update)
DT 01-JUN-2000 (TREMBLrel: 14, Last annotation update)
DE IMMOBILIZATION ANTIGEN PRECURSOR (FRAGMENT).
OS Ichthyophthirius multifiliis.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Opisthokonta; Ichthyophthirius.
OX NCBI_TaxID=5932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GEORGIA;
RX MEDLINE=92335298; PubMed=1631132;
RA Clark T.G., McGraw R.A., Dickerson H.W.;RT "Developmental expression of surface antigen genes in the parasitic
RT ciliate Ichthyophthirius multifiliis."
RL Proc. Natl. Acad. Sci. U.S.A. 89:6363-6367(1992).
RL [12]
RP SEQUENCE FROM N.A.
RC STRAIN=GEORGIA;
RA Clark T.;RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
RN [14]
RP SEQUENCE FROM N.A.
RC STRAIN=GEORGIA;
RA Clark T.;RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL, M9207; AAC36158.1; .
KW Signal.
FT NON_TER
FT SIGNAL
POTENTIAL.

SO	SEQUENCE	394	AA:	39495	MM:	3013C2B2BEFDB682	CRC64:	
	Query Match			100.0%;	Score 94;	DB 5;	Length 394;	
	Best Local Similarity			100.0%;	Pred. NO. 2.3e-07;			
	Matches 15;	Conservative	0;	Mismatches	0;	Indels	0;	
OY	1	CVCKCKPNFYNGSP	15					
DB	107	CVCKCKPNFYNGSP	121					
RESULT	2							
O9XZG2		PRELIMINARY;	PRT;	442	AA.			
ID	O9XZG2							
AC	O9XZG2;							
DT	01-NOV-1999 (TREMBlrel. 12, Created)							
DT	01-NOV-1999 (TREMBlrel. 12, Last sequence update)							
DT	01-JUN-2000 (TREMBlrel. 14, Last annotation update)							
DE	IMMOBILIZATION ANTIGEN PRECURSOR.							
GN	IAG48.							
OS	Ichthyophthirius multifiliis.							
OC	Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;							
OC	Ophryoglenina; Ichthyophthirius.							
OX	NCBI_TaxID=5932;							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RC	STRAIN=G1;							
RX	MEDLINE=99196987; PubMed=10095108;							
RA	Clark T.G., Lin T.L., Jackwood D.A., Sherrill J., Lin Y.,							
RA	Dickerson H.W.;							
RT	"rpe gene for an abundant parasite coat protein predicts tandemly							
RL	repetitive metal binding domains.";							
RL	Gene 229:91-100(1999).							
RP	[2]							
RP	SEQUENCE FROM N.A.							
RC	STRAIN=G1;							
RA	Garltig J., Gao Y., Tishgarten T., Clark T.G., Dickerson H.W.;							
RT	"Surface display of a parasite antigen in the ciliate Tetrahymena							
RT	thermophila.";							
RL	Nat. Biotechnol. 0:0-0(1999).							
DR	EMBL; AF140273; AAD31283.1;							
DR	Signal.							
FT	SIGNAL	1	20	POTENTIAL.				
FT	CHAIN	21	442	IMMOBILIZATION ANTIGEN.				
SO	SEQUENCE	442	AA;	45025	MM;	52658F3F65D27AFA	CRC64;	
	Query Match			100.0%;	Score 94;	DB 5;	Length 442;	
	Best Local Similarity			100.0%;	Pred. NO. 2.5e-07;			
	Matches 15;	Conservative	0;	Mismatches	0;	Indels	0;	
OY	1	CVCKCKPNFYNGSP	15					
DB	127	CVCKCKPNFYNGSP	141					
RESULT	3							
O9NL27		PRELIMINARY;	PRT;	600	AA.			
ID	O9NL27							
AC	O9NL27;							
DT	01-OCT-2000 (TREMBlrel. 15, Created)							
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)							
DT	01-OCT-2000 (TREMBlrel. 15, Last annotation update)							
DE	NETRIN.							
GN	CI-NEP1B.							
OS	Ciona intestinalis.							
OC	Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Phlebobranchia;							
OC	Clonidae; Clona.							
OX	NCBI_TaxID=7719;							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RA	Takamura K.;							

```

RT      "Expression patterns of ascidian netrin homologues."
BL      Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AB033332; BAA94303.1; -.
SQ      SEQUENCE    600 AA;  68312 MW;  227D53324D17CCFB CRC64;

Query Match
Best Local Similarity  52.1%; Score 49; DB 5; Length 600;
Matches      8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      1  CVCKCPNFYYN 11
Db      432  CHKCKPFGFYRN 442

RESULT  4
O9NL28  PRELIMINARY; PRT; 650 AA.
AC      O9NL28;
DT      01-OCT-2000 (TREMBlrel. 15, Created)
DT      01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT      01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE      NETRIN.
GN      CI-NET1A.
OC      Ciona intestinalis.
OC      Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Phlebobranchia;
OX      Cionidae; Ciona.
RN      NCBI_TaxID=7719;
RP      SEQUENCE FROM N.A.
RA      Takamura K.;
RT      "Expression patterns of ascidian netrin homologues."
RL      Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AB033331; BAA94302.1; -.
SQ      SEQUENCE    650 AA;  73807 MW;  F5320C163AFB6E2C CRC64;

Query Match
Best Local Similarity  52.1%; Score 49; DB 5; Length 650;
Matches      8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      1  CVCKCPNFYYN 11
Db      432  CHKCKPFGFYRN 442

RESULT  5
O9NH87  PRELIMINARY; PRT; 401 AA.
AC      O9NH87;
DT      01-OCT-2000 (TREMBlrel. 15, Created)
DT      01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT      01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE      VARIAM-SPECIFIC SURFACE PROTEIN VSP MM/SAC-A/1 (FRAGMENT).
OS      Giardia lamblia (Giardia intestinalis).
OC      Eukaryota; Diplomonadida; Hexamitidae; Giardiae; Giardia.
OX      NCBI_TaxID=5741;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=AD-1;
RA      Mansouri M.; Ey P.L.;
RT      "Analysis of vsp72-like sequences in Giardia intestinalis.";
RL      Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF20275; AAF69829.1; -.
FT      NON_TER
FT      1
SQ      SEQUENCE    401 AA;  40551 MW;  DA973A9CE5120BEF CRC64;

Query Match
Best Local Similarity  51.1%; Score 48; DB 5; Length 401;
Matches      7; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY      1  CVCKCPNFYINGSP 15

```

			:	:	:	:	:
Db	85	CITCKACAGFLKGNP	99				
RESULT	6						
ID	P91904	PRELIMINARY:	PRT: 3704 AA.				
AC	P91904:						
Df	01-MAY-1997 (TREMBLrel. 03, Created)						
Dt	01-MAY-1997 (TREMBLrel. 03, Last sequence update)						
De	01-OCT-2000 (TREMBLrel. 15, Last annotation update)						
DE	LAMININ ALPHA (EPI-1 PROTEIN).						
GN	Epi-1,						
OS	Caenorhabditis elegans.						
OC	Eukaryota; Metazoa; Nematoda; Chromadorea;	Rhabdilitda; Rhabditoidea;					
Ox	Rhabdilitdae; Peloderinae; Caenorrhadtitis.						
NcBI_Taxid=6239;	[1]						
RP	SEQUENCE FROM N.A.						
RC	STRAIN-BRISTOL NZ;						
RA	Joh K., Hedgecock E.M.;						
RN	Submitted (Feb-1997) to the EMBL/GenBank/DDBJ databases.						
RL	(2)						
RM	SEQUENCE FROM N.A.						
RF	STRAIN-BRISTOL NZ;						
RG	Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.						
Berks M.;							
EMBL AB001074;	BAA19229.1;	-					
EMBL AB016806;	BAA32347.1;	-					
EMBL Z70286;	CAB61016.1;	--					
HSSP P02468;	IJKO.						
INTERPRO IPR000034;	-						
INTERPRO IPR000561;	-						
INTERPRO IPR001362;	-						
INTERPRO IPR001542;	-						
INTERPRO IPR001791;	-						
INTERPRO IPR001886;	-						
INTERPRO IPR002049;	-						
PFAM PF00052;	laminin_B_1.						
PFAM PF00053;	laminin_EGF_21.						
PFAM PF00054;	laminin_G_5.						
PFAM PF00055;	laminin_Nterm; 1.						
PRINTS PR00011;	EGFLAMININ.						
PROSITE PS00022;	EGF_1; UNKNOWN_19.						
PROSITE PS00425;	ARTROPOD_DEFENSINS; UNKNOWN_1.						
PROSITE PS00652;	TNFR_NGFR_1; UNKNOWN_1.						
PROSITE PS01186;	EGF_2; 4.						
PROSITE PS01248;	LAMININ_TYPE_EGF; 21.						
PRODOW PD002082;	-; 1.						
PRODOW PD003031;	-; 1.						
Glycoprotein.							
SQ	SEQUENCE 3704 AA; 407842 MW; A2DSB6BDJ153919A CRC64;						
Query Match	Best Local Similarity	51.1%; Score 48; DB 5; Length 3704;					
Matches 7; Conservative	1; Indels	0; Gaps	0;				
DB	685 CDRCKPNFY 693						
QY	1 CVKCKPNFY 9 I : I :						
RESULT	7						
ID	P97939	PRELIMINARY;	PRT: 267 AA.				
AC	P97939:						
Df	01-MAY-1997 (TREMblrel. 03, Created)						
Dt	01-MAY-1997 (TREMblrel. 03, Last sequence update)						

DT	01-OCT-2000 (TREMblrel. 15, last annotation update)
DE	MEMBRANE COFACTOR PROTEIN (FRAGMENT).
CN	MCP.
OC	Cavia porcellus (Guinea pig).
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Hystriiconnath; Cavidae; Cavia.
OX	NCBI_TaxID=10141;
RN	[1]
RN	SEQUENCE FROM N.A.
RC	STRAIN-HARTLEY; TISSUE=LIVER;
RA	Hosokawa M., Nonaka M., Okada N., Nonaka M., Okada H. ;
RT	"Molecular cloning of guinea pig membrane cofactor protein: preferential expression in testis.";
RL	J. Immunol. 0:0-0(1996).
RN	[2]
RN	SEQUENCE FROM N.A.
RC	STRAIN-HARTLEY; TISSUE=LIVER;
RX	MEDLINE=97098723; PubMed=8943400;
RA	Hosokawa M., Nonaka M., Okada N., Nonaka M., Okada H. ;
RT	"Molecular cloning of guinea pig membrane cofactor protein: preferential expression in testis.";
RL	J. Immunol. 157:4946-4952(1996).
DR	EMBL; D84134; BAA12336.1; JOINED.
DR	EMBL; D86340; BAA12236.1; JOINED.
DR	EMBL; D86341; BAA12236.1; JOINED.
DR	EMBL; D86342; BAA12236.1; JOINED.
DR	HSSP; P10998; IYD.
DR	INTERPRO: IPRO00436; -
DR	PfAM: PF00084; sushi. 4.
KW	Membrane.
FT	NON_TER 1
FT	NON_TER 267
SQ	SEQUENCE 267 AA; 29755 MW; 659E2EC15CE3E38C CRC64;
Oy	3 KCKPMFYNGGS 14
Dd	221 QCLPEFFFGSS 232
RESULT	8
P97263	PRELIMINARY: PRT: 337 AA.
AC	P97263.
DT	01-MAY-1997 (TREMblrel. 03, Created)
DT	01-MAY-1997 (TREMblrel. 03, Last sequence update)
DT	01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE	MEMBRANE COFACTOR PROTEIN PRECURSOR.
OS	Cavia porcellus (Guinea pig).
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Hystriiconnath; Cavidae; Cavia.
OX	NCBI_TaxID=10141;
RN	[1]
RN	SEQUENCE FROM N.A.
RC	STRAIN-HARTLEY; TISSUE=TESTIS;
RA	Hosokawa M., Nonaka M., Okada N., Nonaka M., Okada H. ;
RT	"Molecular cloning of guinea pig membrane cofactor protein: preferential expression in testis.";
RL	J. Immunol. 0:0-0(1996).
RN	[2]
RN	SEQUENCE FROM N.A.
RC	STRAIN-HARTLEY; TISSUE=TESTIS;
RX	MEDLINE=97098723; PubMed=8943400;
RA	Hosokawa M., Nonaka M., Okada N., Nonaka M., Okada H. ;
RT	"Molecular cloning of guinea pig membrane cofactor protein: preferential expression in testis.";
RL	J. Immunol. 157:4946-4952(1996).
DR	EMBL; D84132; BAA12234.1; -

DR HSSP; P10998; 1VVD.
 DR INTERPRO; IPR000436; -
 DR PFAM; PF00084; sushi; 4.
 KW Signal; Membrane.
 FT SIGNAL 1 34
 FT CHAIN 35 337
 SO SEQUENCE 337 AA; 37377 MW; 0AFCDFFC4D1FDBB CRC64; POTENTIAL.

Query Match 50.0%; Score 47; DB 11; Length 337;
 Best Local Similarity 58.3%; Pred. No. 6.8;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 CKRPNFYNGGS 14
 : | | | | | | |
 Db 255 QCLPGFYNGSS 266

RESULT 9
 P97261 PRELIMINARY; PRT; 339 AA.
 ID P97261;
 AC P97261;
 DT 01-MAY-1997 (TREMblrel. 03, Created)
 DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE MEMBRANE COFACTOR PROTEIN PRECURSOR (FRAGMENT).
 OS Cavia porcellus (Guinea pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Hystriognathi; Cavidae; Cavia.
 NCBI_Taxid=10141;
 RX STRAIN=HARTLEY; TISSUE=TESTIS;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HARTLEY; TISSUE=TESTIS;
 RA Hosokawa M., Nonaka M., Okada N., Nonaka M., Okada H.;
 RT "Molecular cloning of guinea pig membrane cofactor
 RL protein: preferential expression in testis.";
 RL J. Immunol. 0:0-0(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HARTLEY; TISSUE=TESTIS;
 RX MEDLINE=97098723; PubMed=8943400;
 RA Hosokawa M., Nonaka M., Okada N., Nonaka M., Okada H.;
 RT "Molecular cloning of guinea pig membrane cofactor protein:
 RT preferential expression in testis.";
 RT J. Immunol. 157:4946-4952(1996).
 DR EMBL; D84130; BAA12232.1; -
 DR HSSP; P10998; 1VVD.
 DR INTERPRO; IPR000436; -
 DR PFAM; PF00084; sushi; 4.
 KW Signal; Membrane.
 FT NON_TER 1 1
 FT SIGNAL <1 19
 FT CHAIN 20 339
 SO SEQUENCE 339 AA; 37870 MW; 1461C309F76F02CF CRC64; POTENTIAL.

Query Match 50.0%; Score 47; DB 11; Length 339;
 Best Local Similarity 58.3%; Pred. No. 6.9;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 CKRPNFYNGGS 14
 : | | | | | | |
 Db 240 QCLPGFYNGSS 251

RESULT 10
 ID P70105 PRELIMINARY; PRT; 379 AA.
 AC P70105;
 DT 01-FEB-1997 (TREMblrel. 02, Created)
 DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE MEMBRANE COFACTOR PROTEIN(GMPI-FULL) PRECURSOR.

OS Cavia porcellus (Guinea pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Hystriognathi; Cavidae; Cavia.
 NCBI_Taxid=10141;
 RX STRAIN=HARTLEY; TISSUE=TESTIS;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HARTLEY; TISSUE=TESTIS;
 RA Hosokawa M., Nonaka M., Okada N., Okada H.;
 RT "Molecular cloning of guinea pig membrane cofactor protein:
 RT preferential expression in testis.";
 RT J. Immunol. 0:0-0(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HARTLEY; TISSUE=TESTIS;
 RX MEDLINE=97098723; PubMed=8943400;
 RA Hosokawa M., Nonaka M., Okada N., Nonaka M., Okada H.;
 RT "Molecular cloning of guinea pig membrane cofactor protein:
 RT preferential expression in testis.";
 RT J. Immunol. 157:4946-4952(1996).
 DR EMBL; D88358; BAA13594.1; -
 DR HSSP; P10998; 1VVD.
 DR INTERPRO; IPR000436; -
 DR PFAM; PF00084; sushi; 4.
 KW Signal; Membrane.
 FT SIGNAL 1 34
 FT CHAIN 35 379
 SO SEQUENCE 379 AA; 42159 MW; 5DA56B058CA1573 CRC64; POTENTIAL.

Query Match 50.0%; Score 47; DB 11; Length 379;
 Best Local Similarity 58.3%; Pred. No. 7.7;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 CKRPNFYNGGS 14
 : | | | | | | |
 Db 255 QCLPGFYNGSS 266

RESULT 11
 Q9XVB2 PRELIMINARY; PRT; 335 AA.
 ID Q9XVB2;
 AC Q9XVB2;
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
 DE F01D5.7 PROTEIN.
 GN F01D5.7.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
 OC Rhabdilitidae; Pelodermidae; Caenorhabditis.
 NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Lloyd C.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copesey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smaildon N., Smith A., Sonhammer E., Straden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sprat J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";

RL Nature 368:32-38(1994).
 DR EMBL: Z81493; CAB04038.1; -
 DR INTERPRO: IPR000379; -
 DR INTERPRO: IPR000734; -
 DR PROSITE: PS00120; LIPASE_SER; UNKNOWN_1
 SO SEQUENCE 335 AA; 37402 MW; B813FB06595149D CRC64;

Query Match 47.9%; Score 45; DB 5; Length 335;
 Best Local Similarity 77.8%; Pred. No. 14;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CVCKPNEYNG 9
 |||||
 Db 104 CVCKPNEYNG 112

RESULT 12
 OY1A5 PRELIMINARY; PRT; 540 AA.

AC OY1A5;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE FORIN1 (FRAGMENT).
 GN FURL.
 OS Lymnaea stagnalis (Great pond snail).
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;
 OC Lymnaeidae; Lymnaea.
 OX NCBI_TaxID=6523;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-CEREBRAL GANGLIA;
 RA Spilker S., Smit A.B., Sharp-Baker H.E., Van Elk R.,
 RA Van Kesteren E.R., Van Minnen J., Kurosky A., Gererets W.P.M.;
 RT "The Family of Prohormone Convertases in Lymnaea: Characterization of
 RT Two Alternately Spliced Furin-Like Transcripts and Cell-Specific
 RT Regulation of Their Expression."
 RL J. Neurobiol. 0:0-0(1999).
 DR EMBL: AF140362; AAD4729.1; -
 DR INTERPRO: IPR000209; -
 DR INTERPRO: IPR000294; -
 DR INTERPRO: IPR002884; -
 DR PFAM: PF00082; Peptidase_S8; 1.
 DR PFAM: PF01483; P; 1.
 DR PROSITE: PS00011; GLU_CARBOXYLATION; UNKNOWN_1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 FT NON_TER
 SO SEQUENCE 540 AA; 58790 MW; F6F355699EC399F8 CRC64;

Query Match 47.9%; Score 45; DB 5; Length 540;
 Best Local Similarity 50.0%; Pred. No. 23;
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 CVCKPNEYNG 12
 |||||
 Db 348 CIECKPGEFKMG 359

RESULT 13
 OY1A6 PRELIMINARY; PRT; 967 AA.

AC OY1A6;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE FORIN1-X.
 GN FURL.
 OS Lymnaea stagnalis (Great pond snail).
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;
 OC Lymnaeidae; Lymnaea.
 OX NCBI_TaxID=6523;

RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-CEREBRAL GANGLIA;
 RA Spilker S., Smit A.B., Sharp-Baker H.E., Van Elk R.,
 RA Van Kesteren E.R., Van Minnen J., Kurosky A., Gererets W.P.M.;
 RT "The Family of Prohormone Convertases in Lymnaea: Characterization of
 RT Two Alternately Spliced Furin-Like Transcripts and Cell-Specific
 RT Regulation of Their Expression."
 RL J. Neurobiol. 0:0-0(1999).
 DR EMBL: AF140361; AAD4728.1; -
 DR HSSP: O99405; IMPT.
 DR INTERPRO: IPR000209; -
 DR INTERPRO: IPR000294; -
 DR INTERPRO: IPR002884; -
 DR PFAM: PF00082; Peptidase_S8; 1.
 DR PFAM: PF01483; P; 1.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PROSITE: PS00011; GLU_CARBOXYLATION; UNKNOWN_1.
 DR PROSITE: PS00136; SUBTILASE_ASP; 1.
 DR PROSITE: PS00137; SUBTILASE_HIS; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 SO SEQUENCE 967 AA; 107600 MW; 85EA0988D1B964A4 CRC64;

Query Match 47.9%; Score 45; DB 5; Length 967;
 Best Local Similarity 50.0%; Pred. No. 41;
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 CVCKPNEYNG 12
 |||||
 Db 775 CIECKPGEFKMG 786

RESULT 14
 OY1A6 PRELIMINARY; PRT; 1571 AA.
 AC OY1A6;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE POLYPROTEIN.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BATUMI L;
 RA Tulin A.V., Kogan G.L.;
 RT "New retrotransposon of the EMBL/Genbank/DBJ databases."
 RL Submitted (AUG-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AJ010298; CA09069.1; -
 DR INTERPRO: IPR001584; -
 DR INTERPRO: IPR001969; -
 DR PFAM: PF00665; IVE; 1.
 DR PROSITE: PS00141; ASP_PROTEASE; UNKNOWN_1.
 KW Polyprotein
 SO SEQUENCE 1571 AA; 177711 MW; 5D6492F78F204E67 CRC64;

Query Match 47.9%; Score 45; DB 5; Length 1571;
 Best Local Similarity 77.8%; Pred. No. 66;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 CKPNEYNG 12
 |||||
 Db 463 CKPNEYNG 471

RESULT 15
 OY1A1 PRELIMINARY; PRT; 342 AA.

AC Q9YMY1;
 DT 01-MAY-1999 (TReMBLrel. 10, Created)
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)
 DE LDORF-23 PEPTIDE.
 OS Lymantria dispar multicausid nuclear polyhedrosis virus (LdMNPV).
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 OC Nucleopolyhedrovirus.
 OX NCBI_TaxID=10449;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99124785; Pubmed=9887315;
 RA Kuzio J., Pearson M.N., Harwood S.H., Funk C.J., Evans J.T.,
 RA Slavicek J.M., Rohrmann G.F.;
 RT "Sequence and analysis of the genome of a baculovirus pathogenic for
 RT Lymantria dispar."
 RL Virology 253:17-34(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Kuzio J., Pearson M.N., Harwood S.H., Funk C.J., Evans J.T.,
 RA Slavicek J., Rohrmann G.F.;
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Kuzio J.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF081810; AAC70208.1;
 SQ SEQUENCE 342 AA; 40449 MW; 7EA724D84228BF67 CRC64;

Query Match 46.8%; Score 44; DB 12; Length 342;
 Best Local Similarity 54.3%; Pred. NO. 21;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 CVKCKPNFYNN 11
 1::11 111
 Db 271 CIRCKSRFYKN 281

Search completed: March 6, 2001, 12:52:55
 Job time: 295 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:49:35 ; Search time 83.05 Seconds
(without alignments)
2.470 Million cell updates/sec

Title: US-09-196-161d-7

Perfect score: 32
Sequence: 1 GAPGV 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

_A_Geneseq_36:*

1: /SIDSL/gcgdata/geneseq/geneseqp/AA1980.DAT:*

2: /SIDSL/gcgdata/geneseq/geneseqp/AA1981.DAT:*

3: /SIDSL/gcgdata/geneseq/geneseqp/AA1982.DAT:*

4: /SIDSL/gcgdata/geneseq/geneseqp/AA1983.DAT:*

5: /SIDSL/gcgdata/geneseq/geneseqp/AA1984.DAT:*

6: /SIDSL/gcgdata/geneseq/geneseqp/AA1985.DAT:*

7: /SIDSL/gcgdata/geneseq/geneseqp/AA1986.DAT:*

8: /SIDSL/gcgdata/geneseq/geneseqp/AA1987.DAT:*

9: /SIDSL/gcgdata/geneseq/geneseqp/AA1988.DAT:*

10: /SIDSL/gcgdata/geneseq/geneseqp/AA1989.DAT:*

11: /SIDSL/gcgdata/geneseq/geneseqp/AA1990.DAT:*

12: /SIDSL/gcgdata/geneseq/geneseqp/AA1991.DAT:*

13: /SIDSL/gcgdata/geneseq/geneseqp/AA1992.DAT:*

14: /SIDSL/gcgdata/geneseq/geneseqp/AA1993.DAT:*

15: /SIDSL/gcgdata/geneseq/geneseqp/AA1994.DAT:*

16: /SIDSL/gcgdata/geneseq/geneseqp/AA1995.DAT:*

17: /SIDSL/gcgdata/geneseq/geneseqp/AA1996.DAT:*

18: /SIDSL/gcgdata/geneseq/geneseqp/AA1997.DAT:*

19: /SIDSL/gcgdata/geneseq/geneseqp/AA1998.DAT:*

20: /SIDSL/gcgdata/geneseq/geneseqp/AA1999.DAT:*

21: /SIDSL/gcgdata/geneseq/geneseqp/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	100.0	543	18	Mouse ETS2 repress
2	32	100.0	2201	21	Human ATP binding
3	29	90.6	34	5	Sequence of the co
4	29	90.6	127	21	MAGE8 polypeptide
5	29	90.6	1182	12	Abelson Related Ge
6	28	87.5	95	20	Porcine circovirus
7	28	87.5	224	18	Novel human protei
8	28	87.5	246	21	Human PRO1555 (UNQ
9	28	87.5	313	20	Human polassium ch
10	28	87.5	313	20	Human pTRK-1 prot
11	28	87.5	313	21	Human pTRK family
12	28	87.5	313	21	KT5, a TWIK family

13	28	87.5	385	21	V92056	Human testes speci
14	28	87.5	548	18	W07700	Human ETS2 repress
15	28	87.5	586	20	V23622	Protein encoded by
16	28	87.5	799	21	V79153	Mouse protein kin
17	28	87.5	873	19	W49031	Human Prt1-like su
18	28	87.5	1178	16	R82656	Human mucosal lymp
19	28	87.5	1183	21	V53127	Human circadian rh
20	28	87.5	1841	18	W22605	Tyactone synthase
21	28	87.5	2273	19	W70398	ATP binding casset
22	28	87.5	4630	18	W19629	Streptomyces venez
23	28	87.5	4630	21	V77177	S. venezuelae vep
24	27	84.4	12	17	R90348	Macrophage-colony
25	27	84.4	101	20	V35590	Chlamydia pneumoni
26	27	84.4	296	17	R87038	Human calcium sign
27	27	84.4	382	17	W00496	Candida albicans s
28	27	84.4	436	21	V74432	Neisseria gonorrhe
29	27	84.4	560	19	W34990	Micrococcilla furves
30	27	84.4	784	18	W34179	Human GC binding P
31	27	84.4	927	8	P70768	Epstein-Barr virus
32	27	84.4	1088	17	R88634	Plasmid pASK75 ope
33	27	84.4	1257	15	R46627	Neurocan core prot
34	27	84.4	1277	15	R52702	Plasmid pASK60-str
35	26	81.2	63	19	V21353	Human HUDP-I mutan
36	26	81.2	67	20	V11676	Human 5' EST sece
37	26	81.2	80	20	V13191	Human secreted pro
38	26	81.2	82	20	V35995	Extended human sec
39	26	81.2	83	20	W78170	Human secreted pro
40	26	81.2	111	20	V31679	Fibrocyte cell a
41	26	81.2	111	20	V31713	GRGDP-containing
42	26	81.2	111	20	V31714	GRGDP-containing
43	26	81.2	111	20	V31716	Human secreted pro
44	26	81.2	142	20	W67919	Human secreted pro
45	26	81.2	143	20	W67918	Human secreted pro

ALIGNMENTS

RESULT 1	ID	W07702	standard; Protein; 543 AA.
XX	XX	W07702:	
AC	XX	06-APR-1997	(first entry)
DT	XX	Mouse ETS2 repressor factor (ERF).	
DE	XX	ETS2 repressor factor; ERF; transcriptional repressor;	
XX	XX	tumour suppressor; tumour; cancer; oncoprotein; therapy.	
KW	XX	Mus sp.	
KX	XX		
OS	XX		
FH	XX	key	Location/Qualifiers
FT	XX	Domain	21..98
FT	XX	Domain	/Label=DNA-binding_domain
FT	XX	Domain	/note="ets-like DNA binding domain"
FT	XX	Domain	466..525
FT	XX	Domain	/Label=Active_repressor_domain
PN	XX	W09639517-A1.	
XX	XX	12-DEC-1996.	
PD	XX	04-JUN-1996;	96WO-US10177.
PE	XX	05-JUN-1995;	95US-0469412.
PR	XX	(USSH) US DEPT HEALTH & HUMAN SERVICES.	
XX	XX	Athanasios MA, Beal GJ, Blair DG, Fisher RJ, Mavrothalassitis GJ,	
XX	XX	Sgouras D N;	

DR WPI; 1997-043139/04.
 DR N-PSDB; T47200.
 XX
 PT New DNA encoding ETS2 repressor factor - useful for reducing
 PT tumourigenicity, esp. oncogene associated tumour cells
 PS
 XX Disclosure; Page 70-72; 101pp; English.
 XX
 CC Murine ETS2 repressor factor (ERF) (M07702) is a member of the ETS
 CC family and acts as a transcriptional repressor in mammalian cells.
 CC Its amino acid sequence was deduced from the murine ERF gene
 CC (747198). Human ERF (see also M07700) has also been identified.
 CC ERF has tumour suppressor activity. Chimeric molecules comprising
 CC the ERF repressor domain in combination with a heterologous
 CC transcription factor having a binding domain can be used to reduce
 CC tumourigenicity associated with inappropriate expression of
 CC transcription factors.
 CC
 XX
 SQ Sequence 543 AA;
 QY
 1 GEPAGV 6
 |||||
 Db 463 gepagv 468
 Query Match 100.0%; Score 32; DB 18; Length 543;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 2
 Y79380
 ID Y79380 standard; Protein; 2201 AA.
 XX
 AC Y79380;
 DT 01-AUG-2000 (first entry)
 XX
 DE Human ATP binding cassette ABCA1 (ABCI) protein.
 XX
 KM ABCA1; ABCI: ATP binding cassette; human; cholesterol;
 KM interleukin-1 beta; transporter; inflammation; septic shock;
 KM rheumatoid arthritis; Tangier disease; hypertriglyceridemia;
 KM splenomegaly; atherosclerosis; lipid disorder; dyslipidemia;
 KM psoriasis; lupus erythematosus; diagnosis; therapy.
 XX
 OS Homo sapiens.
 XX
 PN M0200018912-A2.
 XX
 PD 06-APR-2000.
 XX
 PF 21-SEP-1999; 99WO-EP06991.
 XX
 PR 25-SEP-1998; 98US-0101706.
 XX
 PA (PARB) BAYER AG.
 XX
 PI Schmitz G, Klucken J;
 XX
 DR WPI; 2000-293151/25.
 DR N-PSDB; Z94734.
 XX
 PT Adenosine triphosphate binding proteins useful for identifying agents
 PT for treating atherosclerosis and other inflammatory disorders -
 PS
 XX Claim 6; Page 94-105; 154pp; English.
 XX
 CC The present sequence is that of human ATP binding cassette protein
 CC ABCA1 (ABCI), the human homologue of mouse ABCA1 (94% identity). "
 CC The sequence was deduced from ABCA1 cDNA (see Z94734), identified in
 CC a differential screening of human monocytes for cholesterol-sensitive
 CC genes. The ABCA1 gene maps to chromosome 9q22-31. Dysregulated

CC ABCA1 is the gene locus involved in the HDL deficiency syndrome
 CC Tangier disease, associated with hypertriglyceridemia and
 CC splenomegaly. ABCA1 is also a transporter for interleukin-1 beta,
 CC making the gene a candidate for treatment of inflammatory diseases
 CC such as rheumatoid arthritis and septic shock. The invention
 CC also provides other cholesterol-sensitive ABC genes (see Z94735-63)
 CC that can be used for diagnostic and therapeutic applications,
 CC and for biochemical or cell-based assays to screen for
 CC pharmacologically active compounds useful for the treatment of
 CC lipid disorders, atherosclerosis or other inflammatory diseases
 CC such as psoriasis and lupus erythematosus.
 CC
 XX
 SQ Sequence 2201 AA;
 QY
 1 GEPAGV 6
 |||||
 Db 28 gepagv 33
 Query Match 100.0%; Score 32; DB 21; Length 2201;
 Best Local Similarity 100.0%; Pred. No. 4.7e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 3
 P40838
 ID P40838 standard; Protein; 34 AA.
 XX
 AC P40838;
 DT 03-AUG-1992 (first entry)
 XX
 DE Sequence of the connecting peptide of chinchilla insulin precursor.
 XX
 KM Insulin precursor; connecting peptide; diabetes; hormone.
 XX
 OS Chinchilla.
 XX
 PN US4430266-A.
 XX
 PD 07-FEB-1984.
 XX
 PF 16-FEB-1982; 82US-0349397.
 XX
 PR 16-FEB-1982; 82US-0349397.
 PR 27-MAR-1980; 80US-0134389.
 PR 28-NOV-1980; 80US-0210696.
 XX
 PA (ELIT) ELI LILLY & CO.
 XX
 PI Frank BH;
 XX
 DR WPI; 1984-049032/08.
 XX
 PT Insulin precursor prodn. from linear S-sulphonate and mercaptan -
 PT in single step without separate oxidn.
 XX
 PS Disclosure; Column 4; 8pp; English.
 XX
 CC The inventors claim a method for the prepn. of an insulin precursor
 CC in which the A-chain and B-chain are joined through a connecting
 CC peptide. The connecting peptide joins the A-chain at the amino
 CC group of A-1 to the B-chain at the carboxyl group of B-30. The
 CC method is pref. for the prepn. of human insulin precursor (see
 CC P40829). The SOS of the connecting peptides of a number of species
 CC are given (see P40828, P40830-39).
 CC
 XX
 SQ Sequence 34 AA;
 QY
 Query Match 90.6%; Score 29; DB 5; Length 34;
 Best Local Similarity 83.3%; Pred. No. 25;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEAPGV 6
1:||||
DB 10 geapgv 15

RESULT 4
ID Y83166 standard; Protein: 127 AA.
XX Y83166;
AC Y83166;
XX 24-JUL-2000 (first entry)
DT
XX
DE MAGE8 polypeptide.
XX
XX PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube;
KW uterus; placenta; cancer; major histocompatibility complex; MHC; CTL;
KW cytotoxic T lymphocyte; immune response; antibody; drug delivery;
KW immunconjugate.
XX
OS Homo sapiens.
XX
XX WO200012706-A1.
XX
XX 09-MAR-2000.
XX
XX 31-AUG-1999; 99WO-US20046.
XX
XX 01-SEP-1998; 98US-0098993.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Pastan I, Brinkmann U, Vasmataz G, Lee B;
XX WPI: 2000-237869/20.
XX
XX Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T
PT lymphocyte response and for raising antibodies which can be used to
PT detect the presence of PAGE-4 in cell samples or body tissues
XX
XX Disclosure: Figure 1a; 63pp; English.
XX
XX PAGE-4 is a gene preferentially expressed in normal male and female
XX reproductive tissues e.g. prostate, testis, fallopian tube, uterus
XX and placenta, as well as in prostate cancer, testicular cancer and
XX uterine cancer. This expression pattern makes it a target for
XX diagnosis and for vaccine based therapy of such neoplasms.
XX An isolated PAGE-4 peptide which induces a cytotoxic T
XX lymphocyte response when bound to a major histocompatibility complex
XX (MHC) class I molecule or the isolated PAGE-4 protein can be used in
XX immunogenic compositions to raise a cytotoxic T lymphocyte response
XX against cells expressing PAGE-4 including cancer cells of the
XX prostate, uterus and testis. The nucleic acids encoding PAGE-4 or
XX PAGE-4 peptide fragments can also be used in these compositions.
XX Antibodies against PAGE-4 and its peptide fragments can be used in
XX detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell
XX samples or body tissues. The presence of PAGE-4 in tissues which are
XX not related to reproduction can be indicative of the spread of
XX cancerous reproductive tissue. PAGE-4 can also be used to raise
XX antibodies which are then used as the targeting group of
XX immunconjugates comprising toxins used in therapeutic applications.
XX This has applications for drug delivery systems. The PAGE
XX polypeptide shares sequence similarity with the GAGE and MAGE family
XX of proteins.
XX
XX Sequence 127 AA;

Query Match 90.6%; Score 29; DB 21; Length 127;
Best Local Similarity 83.3%; Pred. No. 97;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEAPGV 6
1:||||
DB 20 geapgl 25

RESULT 5
ID R15157 standard; Protein: 1182 AA.
XX R15157;
AC R15157;
XX 12-FEB-1992 (first entry)
DT
XX
DE Abelson Related Gene, B transcript.
XX
XX Arg; diagnosis; therapy; tumour; abl proto-oncogene.
XX
XX Homo sapiens.
XX
XX US7559029-A.
XX
XX 22-OCT-1991.
XX
XX 22-OCT-1991; 91US-0559029.
XX
XX 30-JUL-1990; 90US-0559029.
XX
XX (USSH) NAT INST OF HEALTH.
XX
XX Krub G, Arronson SA, King CR;
XX
XX WPI: 1991-353425/48.
XX N-PSDB: Q14937.
XX
XX Novel human gene related to abl proto-oncogene - designated
PT "Abelson Related Gene", arg, useful for tumour diagnosis and
PT therapy
XX
XX Disclosure: Fig 5D; 40pp; English.
XX
XX The human gene encoding this protein is closely related to but
XX distinct from the abl proto-oncogene and is a member of the tyrosine
XX kinase encoding family of genes. Arg is expressed as two transcripts.
XX By analogy with c-abl, the alternative 5' arg sequences have been
XX designated A (Q14936) and B and it is assumed that they are joined
XX to the arg second exon.
XX The amino acid sequence is represented as found in the specification.
XX
XX Sequence 1182 AA;

Query Match 90.6%; Score 29; DB 12; Length 1182;
Best Local Similarity 83.3%; Pred. No. 9,5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEAPGV 6
1:||||
DB 9 geapgl 14

RESULT 6
ID Y31795 standard; Protein: 95 AA.
XX Y31795;
AC Y31795;
XX
XX 06-DEC-1999 (first entry)
DT
XX
XX Porcine circovirus genome ORF V2-encoded protein.
DE
XX PCV; circovirus; infection; vaccine.
KW
XX Porcine circovirus.

```

XX  W09945956-A1.
PN  16-SEP-1999.
XX  12-MAR-1999; 99WO-US05485.
XX  13-MAR-1998; 98US-0077890.
XX  (UYGE-) UNIV GEORGIA RES FOUND INC.
PA  Poet SE, Ritchie BW, Niagro FD, Lukert PD;
XX  WPI: 1999-561620/47.
DR  N-PSDB; X87992.
XX  New vaccines against circovirus infections in animals
PT  Claim 11; Page 54; 74pp; English.
XX  The present sequence represents a protein encoded by open reading
CC  frame ORF V2 on the viral strand of the porcine circovirus (PCV)
CC  genome (see X87992). The invention provides novel vaccine
CC  compositions which protect against circoviral infections. The
CC  vaccine comprises a nucleic acid vector composed of a eukaryotic
CC  cis-acting transcription/translation regulatory sequence
CC  functionally linked to a nucleic acid encoding a circovirus
CC  polypeptide. In PCV vaccines, the nucleic acid may be any of the
CC  ORFs encoding a PCV polypeptide (see Y31794-99). Nucleic acid
CC  vectors are also used for the transient expression of one or
CC  more PCV polypeptides in a eukaryotic cell. Immunogenic amounts
CC  of such polypeptides can also be used to prevent a PCV-associated
CC  disease in an animal.
XX  Sequence 95 AA:
SQ
Query Match 87.5%; Score 28; DB 20; Length 95;
Best Local Similarity 83.3%; Pred. NO. 1.1e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GEPGV 6
Db 1 1111
25 gevpgv 30
RESULT 7
W36796
ID W36796 standard; Peptide; 224 AA.
XX
AC W36796;
XX
DT 23-APR-1998 (first entry)
XX
DE Novel human protein, designated WWP3, which contains WW domains.
XX
KW Peptide recognition unit; YAP WW domain binding protein; WBP-1; WBP-2;
KW WW domain; cell signalling; growth regulation; cytoskeleton organisation;
KW targeted drug screening; modulator; WW domain interaction; WWP3.
XX
OS Homo sapiens.
XX
FH Key
FT Domain
FT 149..186
FT /note="WW domain"
XX
PN W09737223-A1.
XX
PD 09-OCT-1997.
XX
PF 03-APR-1997; 97WO-US0547.
XX
PR 03-APR-1996; 96US-0630916.

```

```

XX  (CYTO-) CYTOGEN CORP.
PA  (UYNC-) UNIV NORTH CAROLINA.
XX  Fowlkes DM, Kay BK, Pirozzi G;
XX  WPI: 1997-503234/46.
DR  N-PSDB; T95699.
XX  Identifying cell signalling and growth regulatory polypeptides by
PT  reaction with multivalent recognition complex - polypeptides are
PT  useful in targeted drug selection
XX  Claim 48; Fig 21; 220pp; English.
XX  The present sequence represents a novel protein WWP3. The WWP3 gene was
CC  identified and isolated from human bone marrow and brain cDNA libraries,
CC  using peptides W38103-05. These peptide recognition units are based on
CC  the sequences of WW domain binding domains of the YAP WW domain binding
CC  proteins WBP-1 and WBP-2. The WW domain is a small functional domain
CC  found in a large number of proteins from a variety of species including
CC  humans, nematodes and yeast. Its name is derived from the observation
CC  that two tryptophan residues, one in the amino terminal portion of the
CC  WW domain and one in the carboxyl terminal portion, are conserved. Most
CC  proteins containing WW domains have a function involving cell signalling
CC  and growth regulation or the organisation of the cytoskeleton.
CC  Polypeptides containing a WW domain are identified by treating a
CC  multivalent recognition unit complex that has selective binding affinity
CC  for a WW domain, with many polypeptides and identifying those with
CC  selective affinity for the complex. Proteins containing WW domains are
CC  used for targeted drug screening, i.e. to identify potential modulators
CC  of specific WW domain interactions. The valency of the recognition unit
CC  is important in determining specificity of interaction with WW domains.
CC  In multivalent form specificity is relaxed, but not lost, so proteins
CC  containing WW domains similar, but not identical, to the sequence of the
CC  peptides' target WW can be detected, including new polypeptides.
XX  Sequence 224 AA:
SQ
Query Match 87.5%; Score 28; DB 18; Length 224;
Best Local Similarity 83.3%; Pred. NO. 2.7e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GEPGV 6
Db 1 1111
1 gevpgv 6
RESULT 8
Y99438
ID Y99438 standard; Protein; 246 AA.
XX
AC Y99438;
XX
DT 08-AUG-2000 (first entry)
XX
DE Human PRO1555 (UNQ6/3) amino acid sequence SEQ ID NO:338.
XX
KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW Transmembrane; secretion; Immunoadhesion; pharmaceutical; screening.
XX
OS Homo sapiens.
XX
FH
FT
FT
FT
XX
PN W0200012708-A2.
XX
PD 09-MAR-2000.
XX
PF 01-SEP-1999; 99WO-US20111.
XX
PR 01-SEP-1998; 98US-0098716.
XX
PR 01-SEP-1998; 98US-0098749.
XX
PR 01-SEP-1998; 98US-0098750.

```

PR	02-SEP-1998;	98US-0098803.
PR	02-SEP-1998;	98US-0098821.
PR	02-SEP-1998;	98US-0098843.
PR	09-SEP-1998;	98US-0099536.
PR	09-SEP-1998;	98US-0099596.
PR	09-SEP-1998;	98US-0099602.
PR	09-SEP-1998;	98US-0099642.
PR	10-SEP-1998;	98US-0099741.
PR	10-SEP-1998;	98US-0099754.
PR	10-SEP-1998;	98US-0099763.
PR	10-SEP-1998;	98US-0099792.
PR	10-SEP-1998;	98US-0099808.
PR	10-SEP-1998;	98US-0099812.
PR	10-SEP-1998;	98US-0099815.
PR	10-SEP-1998;	98US-0099816.
PR	15-SEP-1998;	98US-0100385.
PR	15-SEP-1998;	98US-0100388.
PR	15-SEP-1998;	98US-0100390.
PR	16-SEP-1998;	98US-0100584.
PR	16-SEP-1998;	98US-0100627.
PR	16-SEP-1998;	98US-0100661.
PR	16-SEP-1998;	98US-0100662.
PR	16-SEP-1998;	98US-0100664.
PR	17-SEP-1998;	98US-0100683.
PR	17-SEP-1998;	98US-0100710.
PR	17-SEP-1998;	98US-0100711.
PR	17-SEP-1998;	98US-0100919.
PR	17-SEP-1998;	98US-0100930.
PR	18-SEP-1998;	98US-0100848.
PR	18-SEP-1998;	98US-0100849.
PR	18-SEP-1998;	98US-0101014.
PR	18-SEP-1998;	98US-0101068.
PR	18-SEP-1998;	98US-0101071.
PR	22-SEP-1998;	98US-0101279.
PR	23-SEP-1998;	98US-0101471.
PR	23-SEP-1998;	98US-0101472.
PR	23-SEP-1998;	98US-0101474.
PR	23-SEP-1998;	98US-0101475.
PR	23-SEP-1998;	98US-0101476.
PR	23-SEP-1998;	98US-0101477.
PR	23-SEP-1998;	98US-0101479.
PR	24-SEP-1998;	98US-0101738.
PR	24-SEP-1998;	98US-0101741.
PR	24-SEP-1998;	98US-0101743.
PR	24-SEP-1998;	98US-0101915.
PR	24-SEP-1998;	98US-0101916.
PR	29-SEP-1998;	98US-0102207.
PR	29-SEP-1998;	98US-0102240.
PR	29-SEP-1998;	98US-0102307.
PR	29-SEP-1998;	98US-0102330.
PR	29-SEP-1998;	98US-0102331.
PR	30-SEP-1998;	98US-0102484.
PR	30-SEP-1998;	98US-0102487.
PR	30-SEP-1998;	98US-0102570.
PR	30-SEP-1998;	98US-0102571.
PR	01-OCT-1998;	98US-0102684.
PR	01-OCT-1998;	98US-0102687.
PR	02-OCT-1998;	98US-0102965.
PR	06-OCT-1998;	98US-0103258.
PR	06-OCT-1998;	98US-0103449.
PR	07-OCT-1998;	98US-0103314.
PR	07-OCT-1998;	98US-0103315.
PR	07-OCT-1998;	98US-0103328.
PR	07-OCT-1998;	98US-0103395.
PR	07-OCT-1998;	98US-0103396.
PR	08-OCT-1998;	98US-0103401.
PR	08-OCT-1998;	98US-0103633.
PR	08-OCT-1998;	98US-0103678.
PR	08-OCT-1998;	98US-0103679.
PR	08-OCT-1998;	98US-0103711.
PR	14-OCT-1998;	98US-0104257.
PR	20-OCT-1998;	98US-0104987.
PR	20-OCT-1998;	98US-0105000.
PR	20-OCT-1998;	98US-0105002.
PR	21-OCT-1998;	98US-0105104.
PR	22-OCT-1998;	98US-0105169.
PR	22-OCT-1998;	98US-0105266.
PR	26-OCT-1998;	98US-0105693.
PR	26-OCT-1998;	98US-0105694.
PR	27-OCT-1998;	98US-0105807.
PR	27-OCT-1998;	98US-0105881.
PR	27-OCT-1998;	98US-0105882.
PR	28-OCT-1998;	98US-0106062.
PR	28-OCT-1998;	98US-0106063.
PR	28-OCT-1998;	98US-0106070.
PR	28-OCT-1998;	98US-0106072.
PR	28-OCT-1998;	98US-0106073.
PR	28-OCT-1998;	98US-0106178.
PR	29-OCT-1998;	98US-0106248.
PR	29-OCT-1998;	98US-0106384.
PR	29-OCT-1998;	98US-0108500.
PR	30-OCT-1998;	98US-0106464.
PR	03-NOV-1998;	98US-0106856.
PR	03-NOV-1998;	98US-0106902.
PR	03-NOV-1998;	98US-0106905.
PR	03-NOV-1998;	98US-0106919.
PR	03-NOV-1998;	98US-0106932.
PR		

Query Match 87.5%; Score 28; DB 21; Length 246;
 Best Local Similarity 100.0%; Pred. No. 3e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEAPG 5
 |||||
 Db 64 gaps 68

RESULT 9
 Y34132
 ID Y34132 standard; Protein; 313 AA.

XX Y34132;

DT 30-NOV-1999 (first entry)

DE Human potassium channel K-Hnov49.

KW Potassium channel; ataxia; arrhythmia; epilepsy; Bartter's syndrome;
 KW cardiovascular disorder; CNS disorder; renal disorder.

OS Homo sapiens.

PN W09943696-A1.

PD 02-SEP-1999.

PE 22-FEB-1999; 99WO-US03826.

PR 19-JAN-1999; 99US-0116448.

PR 25-FEB-1998; 98US-0076687.

PR 07-AUG-1998; 98US-0095836.

PA (AXYS-) AXYS PHARM INC.

PI Curran ME, Hu P, Miller AP, Rutter M, Wang J;

DR WPI: 1999-527591/44.

DR N-PSDB: Z11914.

PT New nucleic acids encoding mammalian K-Hnov potassium channel
 PT proteins, useful for the diagnosis and treatment of episodic ataxia
 PT with myokymia, cardiac arrhythmia, epilepsy and Bartter's syndrome

PS Claim 3; Page 101-102; 112pp; English.

XX This sequence represents the human K-Hnov49 potassium channel.

CC K-Hnov proteins have a high degree of homology to known potassium
 CC channels and may be alpha subunits, which form the functional channel, or

CC accessory subunits that act to modulate the channel activity. K-Hnov49 is

CC a 4 transmembrane domain, 2 pore domain potassium channel. The gene's

CC chromosomal location is 14q1, determined via PCR chromosomal

CC localisation using primers Z11937 and Z11938. K-Hnov CDNAS

CC were isolated by extension of expressed sequence tags (ESTs) which were

CC related but not identical to known human potassium channels. Potential

CC polymorphisms detected as sequence variants between multiple

CC independent clones. Potassium channels have critical roles in various

CC cell types and biochemical pathways. Defective potassium channels are

CC known to cause four human diseases: episodic ataxia with myokymia;
 CC cardiac arrhythmia (long QT syndrome); epilepsy; and Bartter's syndrome.
 CC As potassium channels are critical components of virtually all cells,
 CC it is likely that abnormal potassium channels are also implicated in
 CC certain renal, cardiovascular and central nervous system (CNS) disorders.
 CC Nucleotides encoding K-Hnov proteins may be used for identifying
 CC homologous or related proteins and the DNA sequences encoding them. They
 CC may be used to produce compositions that modulate the expression and
 CC function of the K-Hnov protein and in studying the biochemical pathways
 CC associated with it. They may also be used for the recombinant production
 CC of K-Hnov protein in fermentation cultures. Additionally, such
 CC nucleotides may be used in gene therapy protocols for the treatment
 CC of diseases associated with abnormal potassium channels.

SQ Sequence 313 AA;

Query Match 87.5%; Score 28; DB 20; Length 313;
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEAPG 5
 |||||
 Db 223 gaps 227

RESULT 10
 Y25116
 ID Y25116 standard; Protein; 313 AA.

XX Y25116;

DT 25-AUG-1999 (first entry)

DE Human hTREK-1 protein.

KW hTREK-2; Twik-1 Related K+ channel-2; vasotropic; antiinflammatory;
 KW analgesic; treatment; gene therapy; inhibitor; detection; diagnosis;
 KW disease susceptibility; cerebral; cardiac; renal; ischemia; brain;
 KW inflammation; pain; mimic; neurotransmitter; hormone; chromosome mapping;
 KW linkage analysis; mutation; immunogen; human.

OS Homo sapiens.

PN EP930364-A1.

PD 21-JUL-1999.

PE 16-JAN-1998; 98EP-0400072.

PR 16-JAN-1998; 98EP-0400072.

PA (SYNO) SYNTELABO.

PI Partiseti M;

DR WPI: 1999-387707/33.

DR N-PSDB: X78383.

PT New human polypeptides useful for diagnosing and treating cerebral
 PT and cardiac ischemias

PS Claim 1; Page 18-20; 21pp; English.

XX This invention describes a novel human Twik-1 Related K+ channel-2
 CC (hTREK-2) polypeptide (I) and its encoding nucleic acid (II) which has
 CC vasotropic, antiinflammatory and analgesic activity. (II) or agonists of
 CC (I) may be used to stimulate production of (I) in vivo to treat patients
 CC requiring enhanced activity or expression of (I). This use of (II)
 CC represents a gene therapy regime. Antagonists of (I), the complement of
 CC (II) used as an antisense construct or a polypeptide competitor of (I)
 CC may be administered to patients to inhibit activity or expression of (I).
 CC Detection of the presence or amount of (I) in a sample from a patient or
 CC detection of mutations in (I) may be used to diagnose or measure
 CC susceptibility to diseases related to altered expression or activity of
 CC (I). The diseases and conditions resulting from altered activity or
 CC expression of (I) which may be treated as above include cerebral,
 CC cardiac and renal ischemias, brain and cardiac diseases, inflammation
 CC and pain. In addition, (I), (II), and agonists and antagonists of (I)
 CC may be used to mimic or antagonize the effects of endogenous
 CC neurotransmitters and hormones. (II) or its fragments may be used as
 CC hybridization probes to isolate full length and genomic CDNAS encoding
 CC (I) or its homologues from cDNA or genomic libraries. (II) may also be
 CC used for chromosome mapping and linkage analysis to identify the
 CC relationship between genes and diseases which have been mapped to the
 CC same chromosome. In addition (II) may be used to identify mutations
 CC associated with diseases by comparing the sequence of (II) between

CC affected and unaffected individuals. (1) or its fragments may be used as
 CC immunogens to produce antibodies against (1). Antibodies to (1) may be
 CC used to isolate or identify clones expressing (1) or to purify (1) by
 CC affinity chromatography. These antibodies may also be used to treat the
 CC above diseases as agonists or antagonists of (1).

XX Sequence 313 AA:

Query Match 87.5%; Score 28; DB 20; Length 313;

Best Local Similarity 100.0%; Pred. No. 3.8e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GAPG 5

Db 223 gapg 227

RESULT 11

ID Y68737 standard; Protein; 313 AA.

AC Y68737;

DT 05-MAY-2000 (first entry)

DE KT4, a TWIK family 2PD potassium channel polypeptide.

XX KT4; TWIK family 2PD potassium channel polypeptide: P-domain;

KW expressed sequence tag; EST: AA604914; ion channel dysfunction;

KW renal disease; musculoskeletal disease; proliferative disease;

KW renal failure; nephrosis; cirrhosis; dysphagia; gastritis; myotonia;

KW muscular dystrophy; atherosclerosis; cancer.

OS Homo sapiens.

PN WO200003687-A2.

PD 27-JAN-2000.

PF 20-JUL-1999; 99WO-US16471.

PR 20-JUL-1998; 98US-0093486.

PR 13-AUG-1998; 98US-0096655.

PA (ELAN-) ELAN PHARM INC.

PI Forsayeth JR, Zhao BB, Chavez RA;

DR WPI: 2000-171196/15.

DR N-PSDB; 246092.

PT Novel human potassium channel polynucleotides and polypeptides used in

PT the diagnosis, prevention and treatment of diseases including renal

PT failure, cirrhosis, muscular dystrophy and cancers -

PS Claim 4; Fig 1A-C; 53pp; English.

XX The present sequence represents a protein, designated KT4, which is a

XX member of the TWIK family 2PD potassium channel polypeptides. These

XX polypeptides contain two potential P-domains and 8 (preferably 4)

XX transmembrane domains. The KT4 cDNA sequence was isolated from a brain

XX cDNA library using degenerate oligonucleotides derived from human

XX expressed sequence tag (EST) AA604914. The polypeptides and

CC dysphagia, gastritis, myotonias, muscular dystrophy, atherosclerosis

CC and cancers.

CC Sequence 313 AA:

Query Match 87.5%; Score 28; DB 21; Length 313;

Best Local Similarity 100.0%; Pred. No. 3.8e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GAPG 5

Db 223 gapg 227

RESULT 12

ID Y68738 standard; Protein; 313 AA.

AC Y68738;

DT 05-MAY-2000 (first entry)

DE KT5, a TWIK family 2PD potassium channel polypeptide.

XX KT5; TWIK family 2PD potassium channel polypeptide: P-domain;

KW expressed sequence tag; EST: AA53124; ion channel dysfunction;

KW renal disease; musculoskeletal disease; proliferative disease;

KW renal failure; nephrosis; cirrhosis; dysphagia; gastritis; myotonia;

KW muscular dystrophy; atherosclerosis; cancer.

OS Homo sapiens.

PN WO200003687-A2.

PD 27-JAN-2000.

PF 20-JUL-1999; 99WO-US16471.

PR 20-JUL-1998; 98US-0093486.

PR 13-AUG-1998; 98US-0096655.

PA (ELAN-) ELAN PHARM INC.

PI Forsayeth JR, Zhao BB, Chavez RA;

DR WPI: 2000-171196/15.

DR N-PSDB; 246094.

PT Novel human potassium channel polynucleotides and polypeptides used in

PT the diagnosis, prevention and treatment of diseases including renal

PT failure, cirrhosis, muscular dystrophy and cancers -

PS Claim 7; Fig 3A-C; 53pp; English.

XX The present sequence represents a protein, designated KT5, which is a

XX member of the TWIK family 2PD potassium channel polypeptides. These

XX polypeptides contain two potential P-domains and 8 (preferably 4)

XX transmembrane domains. The KT5 cDNA sequence was isolated from a brain

XX cDNA library using degenerate oligonucleotides derived from human

XX expressed sequence tag (EST) AA53124. The polypeptides and

XX polynucleotides are used in the diagnosis, prevention and treatment of

XX disease states. The polynucleotides may be used to detect and

XX quantitate expression of TWIK family 2PD potassium channels, and

XX aberrant or mutant forms of the polynucleotide which cause various

XX diseases and disorders. Antisense oligonucleotides may be used to

XX modulate the expression of polynucleotides of the invention. The

XX polypeptides are used for treating diseases and disorders associated

XX with ion channel dysfunction, including renal, musculoskeletal and

XX proliferative diseases, e.g. renal failure, nephrosis, cirrhosis,

XX dysphagia, gastritis, myotonias, muscular dystrophy, atherosclerosis

XX and cancers.

SQ Sequence 313 AA:

Query Match 87.5%; Score 28; DB 21; Length 313;
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEAPG 5
 |||||
 Db 223 gaps 227

RESULT 13

ID Y92056 standard; Protein; 385 AA.

XX Y92056;

XX 01-AUG-2000 (first entry)

XX Human testes specific protease 50.

XX Testes specific protease; TSP50; Methyl Differential Display;

XX Mutant Differential Display; breast cancer.

XX Homo sapiens.

XX W0200018238-A1.

XX 06-APR-2000.

XX 28-SEP-1999; 99WO-US22518.

XX 30-SEP-1998; 98US-0163951.

XX 30-JUN-1999; 99US-0345881.

XX (NSHO-) NORTH SHORE-LONG ISLAND JEWISH RES.

XX Duffy HX;

XX WPI: 2000-292958/25.

XX N-PSDB: A08815, A08816.

XX Nucleic acid encoding human tsp50, or immunogenic fragments of it, used
 PT for detecting mutations and methylation patterns in nucleic acid
 PT samples

PS Claim 2; Page 100-101; 104pp; English.

XX This is a testes specific protease, designated TSP50. The TSP50 gene
 CC was found to be differentially expressed in breast cancer cells. It was
 CC identified using a novel method, named "Methyl- (or Mutant-) Differential
 CC Display". The method is used for detecting and isolating DNA sequences
 CC which are mutated or methylated in one tissue type but not in another.
 CC The method is particularly used for detecting and isolating DNA fragments
 CC that are normally methylated but which, for some reason, are
 CC non-methylated in a small proportion of cells, e.g. TSP50 gene sequences.
 CC The methods can also be used to identify and isolate mutations in genomic
 CC DNA, particularly for identifying mutations which are near the promoters
 CC or coding regions of the genes. The enzymes used in the present
 CC invention mean that the proportion of the genome that is being examined
 CC will be enriched for genetically encoded sequences as well as regulatory
 CC sequences, unlike prior art methods. The method uses a combination of
 CC DNA amplification and hybridization/subtraction techniques so that it has
 CC improved sensitivity, compared to prior art methods.

SQ Sequence 385 AA;

Query Match 87.5%; Score 28; DB 21; Length 385;
 Best Local Similarity 100.0%; Pred. No. 4.7e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEAPG 5
 |||||
 Db 41 gaps 45

RESULT 14

ID W07700 standard; Protein; 548 AA.

XX W07700;

XX 06-APR-1997 (first entry)

XX Human ETS2 repressor factor (ERF).

XX ETS2 repressor factor; ERF; transcriptional repressor;

XX tumour suppressor; tumour; cancer; oncoprotein; therapy.

XX Homo sapiens.

XX Key

XX Domain

XX Location/Qualifiers
 29..106
 /label= "DNA_binding_domain
 /note= "ets-like DNA binding domain"
 472..530
 /label= "Active_repressor_domain
 /note= "Claim 20)"

XX W09639517-A1.

XX 12-DEC-1996.

XX 04-JUN-1996; 96WO-US10177.

XX 05-JUN-1995; 95US-0469412.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Athanasios MA, Beal GJ, Blair DG, Fisher RJ, Mayrothalassitis GJ;

XX Sgouras D N;

XX WPI: 1997-043139/04.

XX N-PSDB: T47198.

XX New DNA encoding ETS2 repressor factor - useful for reducing
 PT tumorigenicity, esp. oncogene associated tumour cells
 PT

PS Claim 1; Page 59-61; 101pp; English.

XX Novel human ETS2 repressor factor (ERF) (W07700) is the first member
 CC of the ETS family to be identified as a transcriptional repressor in
 CC mammalian cells. Its amino acid sequence was deduced from a cDNA
 CC clone (T47198) derived from K562 cells. ERF and alternatively
 CC spliced ERF (see also W07701) show no homology to other known
 CC proteins. The ERF repressor domain in combination with a
 CC heterologous transcription factor having a binding domain can
 CC be used as novel transcriptional repressors to reduce
 CC tumorigenicity associated with inappropriate expression of the
 CC GAL4, NF-kappaB (HIV), MYC (Burkitt lymphoma), Fli-1 (Ewing's
 CC sarcoma) and ESR1 transcription factors.

SQ Sequence 548 AA;

Query Match 87.5%; Score 28; DB 18; Length 548;
 Best Local Similarity 100.0%; Pred. No. 6.8e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEAPG 5
 |||||
 Db 469 gaps 473


```

RESULT 15
ID Y23622 standard; Protein; 586 AA.
XX
AC Y23622;
XX
DT 06-SEP-1999 (first entry)
XX
DE Protein encoded by a reading frame of the porcine circovirus genome.
XX
KW MAP: piglet fatal wasting disease; vaccine; circovirus infection;
XX gene therapy.
XX
OS Porcine circovirus.
XX
FH Key Location/Qualifiers
FT Misc-difference 1..586
FT /note= "all x residues are encoded by stop codons"
XX
XX FR2772047-A1.
XX
XX 11-JUN-1999.
XX
XX PD 05-DEC-1997; 97FR-0015396.
XX
XX PF 05-DEC-1997; 97FR-0015396.
XX
XX PR 05-DEC-1997; 97FR-0015396.
XX
XX (NAVE-) CENT NAT ETUD VETERINAIRES & ALIMENTAIRE.
XX
XX PI Albina E, Arnauld C, Blanchard P, Hutet E, Jestin A;
XX PI Le Cann P;
XX
XX WPI: 1999-360000/31.
XX
XX DR N-PSDB; X85593.
XX
XX PT Nucleotide sequence of porcine circovirus MAP - useful in vaccines
XX PT against MAP circovirus infection and in gene therapy
XX
XX PS Claim 7; Fig 3; 89pp; French.
XX
XX CC The present sequence is encoded by a reading frame of the positive
XX CC strand of a porcine circovirus genome which is associated with MAP.
XX CC MAP is the french acronym for piglet fatal wasting disease. The
XX CC polypeptides can be used to detect anti-MAP antibodies. The antibodies
XX CC can be used to detect MAP antigens. The nucleotide sequences can be used
XX CC as probes or primers for detecting MAP nucleic acids. The nucleotide
XX CC sequences, polypeptides, vectors, (pseudoviral particles, transformed
XX CC cells and compounds selected by the screening assay can be used in
XX CC pharmaceutical compositions. The polypeptides, nucleotide sequences,
XX CC vectors and transformed cells can be used in vaccines against MAP
XX CC circovirus infection. The vectors, (pseudoviral particles and
XX CC transformed cells can be used for gene therapy.
XX
SQ Sequence 586 AA;

```

Query Match 87.5%; Score 28; DB 20; Length 586;
 Best Local Similarity 83.3%; Pred. No. 7.3e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GEARGV 6
 .|||
 Db 412 gevpgv 417

Search completed: March 6, 2001, 12:49:37
 Job time: 98 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:53:50 ; Search time 57.76 Seconds
(without alignments)
1.865 Million cell updates/sec

Title: US-09-196-161D-7

Perfect score: 32

Sequence: 1 GEAPEV 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/PCITUS.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	100.0	543	2 US-08-469-412A-7	Sequence 7, Appli
2	29	90.6	153	3 US-08-851-843A-177	Sequence 177, App
3	28	90.6	153	3 US-08-974-549A-296	Sequence 236, App
4	28	87.5	224	3 US-08-630-916A-50	Sequence 50, Appl
5	28	87.5	548	2 US-08-469-412A-2	Sequence 2, Appli
6	28	87.5	873	3 US-08-990-140-2	Sequence 2, Appli
7	28	87.5	1178	1 US-08-199-776-2	Sequence 2, Appli
8	28	87.5	1178	1 US-08-663-731-2	Sequence 2, Appli
9	28	87.5	1178	3 US-08-879-338-2	Sequence 2, Appli
10	28	87.5	1178	4 PCT-US95-02044-2	Sequence 2, Appli
11	28	87.5	1841	2 US-08-804-227C-6	Sequence 6, Appli
12	27	84.4	12	1 PCT-US95-05471-52	Sequence 52, Appl
13	27	84.4	12	1 PCT-US95-05471-52	Sequence 52, Appl
14	27	84.4	48	1 US-08-340-428B-26	Sequence 26, Appl
15	27	84.4	48	4 PCT-US93-07306-26	Sequence 26, Appl
16	27	84.4	88	2 US-08-751-767A-54	Sequence 54, Appl
17	27	84.4	296	1 US-08-261-662-2	Sequence 2, Appli
18	27	84.4	296	4 PCT-US95-07752-2	Sequence 2, Appli
19	27	84.4	333	1 US-08-225-477B-4	Sequence 4, Appli
20	27	84.4	333	4 PCT-US95-04353-4	Sequence 4, Appli
21	27	84.4	382	2 US-08-360-606B-30	Sequence 30, Appli
22	27	84.4	836	3 US-08-486-099-103	Sequence 103, App
23	27	84.4	856	3 US-08-484-223B-103	Sequence 103, App
24	27	84.4	856	3 US-08-919-597-103	Sequence 103, App
25	27	84.4	856	3 US-08-475-668A-103	Sequence 103, App
26	27	84.4	856	3 US-08-485-551A-103	Sequence 103, App
27	27	84.4	856	3 US-08-471-913A-103	Sequence 103, App
28	27	84.4	857	1 US-08-220-151-10	Sequence 10, Appl

29	27	84.4	857	1 US-08-413-118-10	Sequence 10, Appl
30	27	84.4	857	3 US-08-804-439A-18	Sequence 18, Appl
31	27	84.4	857	3 US-08-360-107A-113	Sequence 113, App
32	27	84.4	857	3 US-08-473-446-10	Sequence 10, Appl
33	27	84.4	857	3 US-08-720-229-18	Sequence 18, Appl
34	27	84.4	1257	1 US-08-340-428B-49	Sequence 49, Appl
35	26	81.2	237	2 US-08-760-745-5	Sequence 5, Appli
36	26	81.2	311	2 US-08-391-916A-6	Sequence 6, Appli
37	26	81.2	383	2 US-08-391-916A-4	Sequence 4, Appli
38	26	81.2	384	1 US-08-350-435-2	Sequence 2, Appli
39	26	81.2	384	4 PCT-US95-15916-2	Sequence 2, Appli
40	26	81.2	404	1 US-08-203-716-2	Sequence 2, Appli
41	26	81.2	404	1 US-08-242-663A-2	Sequence 2, Appli
42	26	81.2	404	1 US-08-440-179-2	Sequence 2, Appli
43	26	81.2	404	2 US-08-450-130A-1	Sequence 1, Appli
44	26	81.2	404	2 US-08-391-916A-2	Sequence 2, Appli
45	26	81.2	404	2 US-08-573-890-2	Sequence 2, Appli

ALIGNMENTS

```

RESULT
1
US-08-469-412A-7
: Sequence 7, Application US/08469412A
: Patent No. 5856125
:
: GENERAL INFORMATION:
: APPLICANT: Mavrothalassitis, George J.
: APPLICANT: Blair, Donald G.
: APPLICANT: Fisher, Robert J.
: APPLICANT: Beal Jr., Gregory J.
: APPLICANT: Athanasiou, Metropi A.
: APPLICANT: Sgouras, Dionysios N.
: TITLE OF INVENTION: The ERF Genetic Locus and Its Products
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/469,412A
: FILING DATE: 05-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Garrett-Wackowski, Eugenia
: REGISTRATION NUMBER: 37,330
: REFERENCE/DOCKET NUMBER: 015280-229000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 543 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: Linear
: MOLECULE TYPE: protein
: FEATURE:
: NAME/KEY: Protein
: LOCATION: 1..543
: OTHER INFORMATION:
: OTHER INFORMATION: (note= "murine ERF amino acid sequence
: OTHER INFORMATION: (first 8 amino acids from first exon not
: included")
:
: US-08-469-412A-7

```

Query Match 100.0%; Score 32; DB 2; Length 543;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEARCV 6
DB 463 GEARCV 468

RESULT 2

US-08-851-843A-177
Sequence 177, Application US/08851843A
Patent No. 6093809
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin H.
APPLICANT: Andrews, William H.
TITLE OF INVENTION: No. 6093809el Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/851,843A
FILING DATE: 06-MAY-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002930US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 177:
SEQUENCE CHARACTERISTICS:
LENGTH: 153 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-851-843A-177

Query Match 90.6%; Score 29; DB 3; Length 153;
Best Local Similarity 83.3%; Pred. No. 83;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEARCV 6
DB 74 GEARCV 79

RESULT 3

US-08-974-549A-296
Sequence 296, Application US/08974549A
Patent No. 6166178
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin H.
APPLICANT: Andrews, William H.
TITLE OF INVENTION: Human Telomerase Catalytic Subunit
NUMBER OF SEQUENCES: 727
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,549A
FILING DATE: 19-NOV-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002610US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 296:

SEQUENCE CHARACTERISTICS:
LENGTH: 153 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-974-349A-296

Query Match 90.6%; Score 29; DB 3; Length 153;
Best Local Similarity 83.3%; Pred. No. 83;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEAPGV 6
|||||
DB 74 GEAPGL 79

RESULT 4
US-08-630-916A-50
Sequence 50, Application US/08630916A
Patent No. 6011137
GENERAL INFORMATION:
APPLICANT: Pirozzi, Gregorio
APPLICANT: Kay, Brian K.
APPLICANT: Fowlkes, Dana M.
TITLE OF INVENTION: IDENTIFICATION AND ISOLATION OF NOVEL
TITLE OF INVENTION: POLYPEPTIDES HAVING WW DOMAINS AND METHODS OF USING SAME
NUMBER OF SEQUENCES: 124
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,916A
FILING DATE: 03-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MISROCK, S. LESLIE
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-203
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 896-8864/9741
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 224 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-630-916A-50

Query Match 87.5%; Score 28; DB 3; Length 224;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GEAPGV 6
|||||
DB 1 GEVPGV 6

RESULT 5
US-08-469-412A-2

Sequence 2, Application US/08469412A
Patent No. 5856125
GENERAL INFORMATION:
APPLICANT: Mavrothalassitis, George J.
APPLICANT: Blair, Donald G.
APPLICANT: Fisher, Robert J.
APPLICANT: Beal Jr., Gregory J.
APPLICANT: Athanasiou, Meropi A.
APPLICANT: Sgouras, Dionysios N.
TITLE OF INVENTION: The ERF Genetic Locus and Its Products
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,412A
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Garrett-Wackowski, Eugenia
REGISTRATION NUMBER: 37,330
REFERENCE/DOCKET NUMBER: 015280-229000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 548 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-469-412A-2

Query Match 87.5%; Score 28; DB 2; Length 548;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEAPG 5
|||||
DB 469 GEAPG 473

RESULT 6
US-08-990-140-2
Sequence 2, Application US/08990140A
Patent No. 6093795
GENERAL INFORMATION:
APPLICANT: Olsen, Henrik S.
APPLICANT: Ruben, Steven M.
APPLICANT: Sonenberg, Nahum
APPLICANT: Methot, Nathalie
APPLICANT: Rom, Fran
TITLE OF INVENTION: Human Prt1-like Subunit Protein (hprt1) and Human
TITLE OF INVENTION: eIF4G1-like Protein (p97) Genes
FILE REFERENCE: 1488.0700001
CURRENT APPLICATION NUMBER: US/08/990,140A
CURRENT FILING DATE: 1997-12-12
EARLIER APPLICATION NUMBER: US 60/033,151
EARLIER FILING DATE: 1996-12-13
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 873

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-990-140-2

Query Match      87.5%; Score 28; DB 3; Length 873;
Best Local Similarity 100.0%; Pred. NO. 7.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GEPG 5
        |||||
Db      105 GEPG 109

RESULT 7
US-08-199-776-2
; Sequence 2, Application US/08199776
; Patent No. 5594120
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael B.
; APPLICANT: Parker, Christina M.
; TITLE OF INVENTION: NO. 5594120e1 Integrin alpha subunit
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield and Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/199,776
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1178 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-199-776-2

Query Match      87.5%; Score 28; DB 1; Length 1178;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GEPG 5
        |||||
Db      89 GEPG 93

RESULT 8
US-08-663-731-2
; Sequence 2, Application US/08663731
; Patent No. 6057423
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael B.
; APPLICANT: Parker, Christina M.
; TITLE OF INVENTION: NO. 6057423e1 Integrin alpha subunit
; NUMBER OF SEQUENCES: 25
```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield and Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/663,731
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/199,776
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1178 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-663-731-2

Query Match      87.5%; Score 28; DB 3; Length 1178;
Best Local Similarity 100.0%; Pred. NO. 1e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GEPG 5
        |||||
Db      89 GEPG 93

RESULT 9
US-08-879-338-2
; Sequence 2, Application US/08879338A
; Patent No. 6063906
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael B.
; APPLICANT: Parker, Christina M.
; TITLE OF INVENTION: Antibodies to No. 6063906e1 Integrin Alpha
; FILE REFERENCE: B0801/7080/ERP
; CURRENT APPLICATION NUMBER: US/08/879,338A
; CURRENT FILING DATE: 1997-06-20
; EARLIER APPLICATION NUMBER: US 08/663,731
; EARLIER FILING DATE: 1996-06-14
; EARLIER APPLICATION NUMBER: US 08/199,776
; EARLIER FILING DATE: 1994-02-18
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1178
; TYPE: PRT
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (-18)...(-1)
US-08-879-338-2

Query Match      87.5%; Score 28; DB 3; Length 1178;
```

Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEAPG 5
|||||
DB 89 GEAPG 93

RESULT 10

PCT-US95-02044-2
Sequence 2, Application PC/TUS9502044
GENERAL INFORMATION:

APPLICANT:
TITLE OF INVENTION: Novel integrin alpha subunit
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield and Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02210

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/02044
FILING DATE: herewith

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/199,776
FILING DATE: 18 February 1994
ATTORNEY/AGENT INFORMATION:

NAME: plumer, Elizabeth R.
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7020
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1178 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
PCT-US95-02044-2

Query Match 87.5%; Score 28; DB 4; Length 1178;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEAPG 5
|||||
DB 89 GEAPG 93

RESULT 11

US-08-804-227C-6
Sequence 6, Application US/08804227C
Patent No. 5876991
GENERAL INFORMATION:

APPLICANT: Dehoff, Bradley S.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Kosteck, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER

CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285

COMPUTER READABLE FORM:

COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231

TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1841 amino acids
TYPE: amino acid
TOPOLOGY: unknown

MOLECULE TYPE: protein
US-08-804-227C-6

Query Match 87.5%; Score 28; DB 2; Length 1841;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEAPG 5
|||||
DB 474 GEAPG 478

RESULT 12

US-08-260-582-52
Sequence 52, Application US/08260582
Patent No. 5635182
GENERAL INFORMATION:

APPLICANT: McCoy, John M.
APPLICANT: Lu, Zhijian
TITLE OF INVENTION: METHOD OF DETECTING LIGAND
NUMBER OF SEQUENCES: 76
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.

ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/260,582
FILING DATE: 16-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Weinert, M. C.
REGISTRATION NUMBER: 31,544
REFERENCE/DOCKET NUMBER: GI 5236

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-260-582-52

Query Match 84.4%; Score 27; DB 1; Length 12;
Best Local Similarity 83.3%; Pred. No. 15;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GEARCV 6
11111
Db 5 GELPGV 10

RESULT 13
PCT-US95-05471-52
Sequence 52, Application PC/TUS9505471
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: METHOD OF DETECTING LIGAND INTERACTIONS
NUMBER OF SEQUENCES: 76
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05471
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
PCT-US95-05471-52

Query Match 84.4%; Score 27; DB 4; Length 12;
Best Local Similarity 83.3%; Pred. No. 15;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GEARCV 6
11111
Db 5 GELPGV 10

RESULT 14
US-08-340-428B-26
Sequence 26, Application US/08340428B
Patent No. 5648465
GENERAL INFORMATION:
APPLICANT: MARGOLIS, Richard U.
APPLICANT: RAUCH, Uwe
APPLICANT: MARGOLIS, Renee K.
TITLE OF INVENTION: CLONING, EXPRESSION AND USES FOR A
TITLE OF INVENTION: NEUROCAN AS A CHONDROITIN SULFATE PROTEOGLYCAN
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/340,428B
FILING DATE: 14 No. 5648465ember 1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/922,911
FILING DATE: 03 August 1992
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: Margolis-1A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 48 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-340-428B-26

Query Match 84.4%; Score 27; DB 1; Length 48;
Best Local Similarity 83.3%; Pred. No. 61;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GEARCV 6
11111
Db 19 GSAPCV 24

RESULT 15
PCT-US93-07306-26
Sequence 26, Application PC/TUS9307306
GENERAL INFORMATION:
APPLICANT: MARGOLIS, Richard U.
APPLICANT: RAUCH, Uwe
APPLICANT: MARGOLIS, Renee K.
TITLE OF INVENTION: CLONING, EXPRESSION AND USES FOR A
TITLE OF INVENTION: NEUROCAN AS A CHONDROITIN SULFATE PROTEOGLYCAN
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07306
FILING DATE: 03-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/922,911
FILING DATE: 03-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, Guy K.
REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: Margolis-1A PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:
LENGTH: 48 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US93-07306-26

Query Match 84.4%; Score 27; DB 4; Length 48;
Best Local Similarity 83.3%; Pred. No. 61;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 1 GEAPGV 6
| | | | |
Db 19 GSAPGV 24

Search completed: March 6, 2001, 12:53:51
Job time: 351 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:50:54 ; Search time 70.34 Seconds
(without alignments)
5.792 Million cell updates/sec

Title: us-09-196-161d-7

Perfect score: 32

Sequence: 1 GEPGV 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.66:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	100.0	395	2 A46031	Immobilization sur
2	32	100.0	1010	2 140329	brkA proteoin - Bor
3	32	100.0	2201	2 A54774	App binding casset
4	31	96.9	170	2 S11918	pulh protein - Kle
5	31	96.9	1340	2 A39808	proteoglycan core
6	31	96.9	2327	2 T42630	aggrecaan - bovine
7	31	96.9	26926	1 I38344	titin, cardiac mus
8	29	90.6	234	2 I38667	melanoma antigen M
9	29	90.6	301	2 C82615	haloalkane dehalog
10	29	90.6	321	1 S43131	thioredoxin reduct
11	29	90.6	402	2 F83189	probable aminotran
12	29	90.6	413	2 JC5178	probable stavalio
13	29	90.6	424	2 H83347	hypothetical prote
14	29	90.6	443	2 T35776	hypothetical prote
15	29	90.6	526	2 F83268	probable FAD-depen
16	29	90.6	577	2 S33763	hormone receptor N
17	29	90.6	732	2 D64754	probable oxidoredu
18	29	90.6	1040	2 E71412	hypothetical prote
19	29	90.6	1182	2 A35662	protein-tyrosine k
20	29	90.6	2109	2 I50421	aggrecaan precursor
21	28	87.5	104	2 A72649	hypothetical prote
22	28	87.5	114	2 T44547	hypothetical prote
23	28	87.5	114	2 F83567	hypothetical prote
24	28	87.5	130	2 H72642	hypothetical prote
25	28	87.5	137	2 T49243	hypothetical prote
26	28	87.5	146	2 T49867	hypothetical prote
27	28	87.5	157	2 E75373	hypothetical prote
28	28	87.5	162	2 C72699	hypothetical prote
29	28	87.5	166	2 E72607	hypothetical prote

30	28	87.5	177	2 T42289	hypothetical prote
31	28	87.5	185	2 T00519	proline-rich prote
32	28	87.5	195	2 E70446	N-terminus of phag
33	28	87.5	196	2 E75111	hypothetical prote
34	28	87.5	197	1 C71022	hypothetical prote
35	28	87.5	199	2 E69534	DNA polymerase, ba
36	28	87.5	200	2 S34604	prolactin - marble
37	28	87.5	215	2 E72736	probable DNA polym
38	28	87.5	221	2 B48266	protein-tyrosine k
39	28	87.5	225	2 T35127	hypothetical prote
40	28	87.5	258	2 JC6327	4-hydroxy-2-oxoval
41	28	87.5	258	2 E71646	hypothetical prote
42	28	87.5	259	2 G70635	hypothetical prote
43	28	87.5	277	2 C71351	probable DNA polym
44	28	87.5	284	2 JC5687	signal peptidase I
45	28	87.5	295	2 E75366	glutamyl-tRNA synt

ALIGNMENTS

RESULT 1

A46031

Immobilization surface T-antigen precursor - Ichthyophthirius multifiliis (fragment)

C:Species: Ichthyophthirius multifiliis

C>Date: 29-Sep-1999 #sequence_revision 29-Sep-1999 #text_change 07-Dec-1999

C:Accession: A46031

R:Clark, T.G.; McGraw, R.A.; Dickerson, H.W.

A>Title: Developmental expression of surface antigen genes in the parasitic ciliate I

A:Reference number: A46031; MUID:92335298

A:Accession: A46031

A:Molecule type: mRNA; protein

A:Residues: 1-395 <C1A>

A:Cross-references: GB:M82907; MID:93628568; PIDN:ANC36158.1; PID:93628569

A>Note: the authors translated the codon UUG for residue 330 as Ile

A>Note: sequence extracted from NCBI backbone (NCBIN:108734, NCBIPI:108735); the seque

C:Genetics:

A:Genetic code: SGC5

C:Keywords: glycoprotein; surface antigen

F:2-395/Product: immobilization surface T-antigen #status experimental <MAY>

F:156,191,245,281/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 32; DB 2; Length 395;

Best Local Similarity 100.0%; Pred. No. 46;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEPGV 6

DB 124 GEPGV 129

RESULT 2

140329

brkA proteoin - Bordetella pertussis

C:Species: Bordetella pertussis

C>Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999

C:Accession: 140329

R:Fernandez, R.C.; Weiss, A.A.

A>Title: Cloning and sequencing of a Bordetella pertussis serum resistance locus.

A:Reference number: 140328; MUID:95012680

A:Accession: 140329

A>Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-1010 <RES>

A:Cross-references: EMBL:U12276; MID:9562025; PIDN:AAA51646.1; PID:9562026

C:Genetics:

A:Gene: brkA

Query Match 100.0%; Score 32; DB 2; Length 1010;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEARCV 6
|||||
Db 185 GEARCV 190

RESULT 3

A54774
ATP binding cassette transporter ABC1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 05-Apr-1995 #sequence_revision 05-Apr-1995 #text_change 17-Mar-2000
C:Accession: A54774
R:Luciani, M.F.; Denizot, F.; Savary, S.; Mattel, M.G.; Chimini, G.
Genomics 21, 150-159, 1994
A:Title: Cloning of two novel ABC transporters mapping on human chromosome 9.
A:Reference number: A54774; MUID:94375008
A:Accession: A54774
A:Molecule type: mRNA
A:Residues: 1-2201 <LNC>
A:Cross-references: GB:X75926; NID:g495256; PIDN:CA53530.1; PID:g495257
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
C:Keywords: ATP; duplication; P-loop
F:856-1047/Domain: ATP-binding cassette homology <ABC1>
F:873-880/Region: nucleotide-binding motif A (P-loop)
F:1869-2060/Domain: ATP-binding cassette homology <ABC2>
F:1886-1893/Region: nucleotide-binding motif A (P-loop)

Query Match 100.0%; Score 32; DB 2; Length 2201;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEARCV 6
|||||
Db 28 GEARCV 33

RESULT 4

S11918
puli protein - Klebsiella oxytoca
C:Species: Klebsiella oxytoca
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Nov-1999
C:Accession: S11918
R:Reyess, I.; Pugsley, A.P.
Mol. Gen. Genet. 222, 176-184, 1990
A:Title: Five additional genes in the pulC-O operon of the gram-negative bacterium Klebs
A:Reference number: S11917; MUID:91109698
A:Accession: S11918
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-170 <REY>
C:Superfamily: secretion protein xcpu

Query Match 96.9%; Score 31; DB 2; Length 170;
Best Local Similarity 83.3%; Pred. No. 33;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEARCV 6
|||||
Db 149 GEARCV 154

RESULT 5

A39808
proteoglycan core protein, cartilage - bovine (fragments)
N:Alternate names: aggrecan; aggregating cartilage proteoglycan
C:Species: Bos primigenius taurus (cattle)
C:Date: 20-Mar-1992 #sequence_revision 23-Mar-1995 #text_change 13-Aug-1999
C:Accession: A34234; A27752; A39808; A27751; E29164; B27751; D27751; E27751; F27
R:Antonsson, P.; Heinegard, D.; Oldberg, A.

J. Biol. Chem. 264, 16170-16173, 1989
A:Title: The keratan sulfate-enriched region of bovine cartilage proteoglycan consist
A:Reference number: A34234; MUID:89380219
A:Accession: A34234
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 128-621 <ANT>
A:Cross-references: GB:J05028
R:Oldberg, A.; Antonsson, P.; Heinegard, D.
Biochem. J. 243, 255-259, 1987

A:Title: The partial amino acid sequence of bovine cartilage proteoglycan, deduced fro
A:Reference number: A27752; MUID:87270630
A:Accession: A27752
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 622-1340 <OLD>
R:Sandy, J.D.; Boynton, R.E.; Flannery, C.R.
J. Biol. Chem. 266, 8198-8205, 1991

A:Title: Analysis of the catabolism of aggrecan in cartilage explants by quantitation
A:Reference number: A39808; MUID:91217051
A:Accession: A39808
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-28;59-82;131-137, 'QSET', 142-149;196-207;226-249;1137-1143;1252-1267;127
R:Perin, J.P.; Bonnet, F.; Jolles, P.
FEBS Lett. 206, 73-77, 1986

A:Title: Structural relationship between link proteins and proteoglycan monomers.
A:Reference number: A27751; MUID:87005253
A:Accession: A27751
A:Molecule type: protein
A:Residues: 29-58;74-130;174-175, 'A', 177-204;208-225 <PER>
R:Perin, J.P.; Bonnet, F.; Jolles, P.
FEBS Lett. 176, 37-42, 1984

A:Title: Sequence data concerning the protein core of the cartilage proteoglycan mono
A:Reference number: A91327; MUID:85027710
A:Accession: E29164
A:Molecule type: protein
A:Residues: 1230-1249 <PE2>
C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology;
C:Keywords: glycoprotein
F:8-28/Domain: link protein repeat homology (fragment) <LNK1>
F:29-58/Domain: link protein repeat homology (fragment) <LNK2>
F:80-146/Domain: link protein repeat homology (fragment) <LNK3>
F:167-248/Domain: link protein repeat homology <LNK4>
F:1130-1250/Domain: C-type lectin homology <LCH>
F:1257-1313/Domain: complement factor H repeat homology <RHD>

Query Match 96.9%; Score 31; DB 2; Length 1340;
Best Local Similarity 83.3%; Pred. No. 2.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEARCV 6
|||||
Db 637 GEARCV 642

RESULT 6

T42630
aggrecan - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000
C:Accession: T42630
R:Herzig, T.M.; Kollar, J.; Huynh, T.D.
submitted to the EMBL Data Library, September 1996
A:Description: Complete coding sequence of bovine aggrecan: comparative structural an
A:Reference number: 222182
A:Accession: T42630
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2327 <HER>
A:Cross-references: EMBL:U76615; NID:91730259; PID:91730260; PIDN:AAB38524.1
A:Experimental source: articular chondrocytes

C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; EG
C:Keywords: cartilage; chondroitin sulfate proteoglycan; extracellular matrix; glycoprot

Query Match 96.9%; Score 31; DB 2; Length 2327;
Best Local Similarity 83.3%; Pred. No. 3.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEPGV 6
Db 1624 GEPGCI 1629

RESULT 7
138344
titin, cardiac muscle [validated] - human
N:Alternate names: connectin
N:Contains: serine/threonine-specific protein kinase (EC 2.7.1.-)
C:Species: Homo sapiens (man)
C:Date: 12-Aug-1996 #sequence, revision 12-Aug-1996 #text, change 15-Sep-2000
C:Accession: 138344; 138345; S20898; S20897; S20899; S63665; S37393
R:Labelt, S.; Kolmerer, B.
Science 270, 293-296, 1995
A:Title: titins: giant proteins in charge of muscle ultrastructure and elasticity.
A:Reference number: A57430; MUID:96026330
A:Accession: 138344
A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL
A:Molecule type: mRNA
A:Residues: 1-26926 <LAB1>
A:Cross-references: EMBL:X90568; NID:91017424; PID:91017425
R:Musco, G.; Tziatzios, C.; Schuck, P.; Pastore, A.
Biochemistry 34, 553-561, 1995
A:Title: Dissecting titin into its structural motifs: identification of an alpha-helix
A:Reference number: 138345; MUID:95119041
A:Accession: 138345
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1977-2014 <MUS>
A:Cross-references: EMBL:X83270; NID:9602579; PIDN:CAA58243.1; PID:9602580
R:Labelt, S.; Gautel, M.; Lakey, A.; Trnlick, J.
EMBO J. 11, 1711-1716, 1992
A:Title: Towards a molecular understanding of titin.
A:Reference number: S20897; MUID:92258380
A:Accession: S20898
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 16330-16382, 'S', 16384-16756, 'F', 16758-16860 <LAB2>
A:Cross-references: EMBL:X64699; NID:937190; PIDN:CAA45939.1; PID:937193
A:Accession: S20897
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 16330-16382, 'S', 16384-16756, 'F', 16758-16860 <LAB2>
A:Cross-references: EMBL:X64699; NID:937190; PIDN:CAA45940.1; PID:937191
A:Accession: S20899
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 'P', 22278-22431, 'R', 22433-22448, 'G', 22450-22453, 'Q', 22455-22480, 'TR', 22483-2
A:Cross-references: EMBL:X64699; NID:937190; PIDN:CAA45938.1; PID:937195
R:Kolmerer, B.; Oliveri, N.; Wilt, C.C.; Herrmann, B.G.; Labelt, S.
J. Mol. Biol. 256, 556-563, 1996
A:Title: Genomic organization of M line titin and its tissue-specific expression in two
A:Reference number: S63665; MUID:96177761
A:Accession: S63665
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 26729-26825 <KOL>
A:Cross-references: EMBL:X92412; NID:91236761
R:Gautel, M.; Leonard, K.; Labelt, S.
EMBO J. 12, 3827-3834, 1993
A:Title: Phosphorylation of KSP motifs in the C-terminal region of titin in differentiat
A:Reference number: S37393; MUID:94008990
A:Accession: S37393

A:Molecule type: mRNA
A:Residues: 26831-26926 <GAU>
R:Improta, S.; Pollou, A.S.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, February 1996
A:Reference number: A66736; PDB:1TTT
A:Contents: annotation: conformation by (1)H-NMR, residues 5253-5341
R:Pinhl, M.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, August 1996
A:Reference number: A66201; PDB:1NCT
A:Contents: annotation: conformation by (1)H-NMR, residues 'S', 26059-26155
C:Genetics:
A:Gene: GDB:TTN
A:Cross-references: GDB:127867; OMIM:188840
A:Map position: 2q31-2q32
C:Function:
A:Description: structural protein forming filaments in striated muscle
C:Superfamily: titin, fibronectin type III repeat homology; immunoglobulin homology;
C:Keywords: alternative splicing; calmodulin binding; cardiac muscle; duplication; gl
structural protein
F:24752-25008/Domain: protein kinase homology <KIN>
F:84,177,905,2276,2378,2459,2481,2563,2669,2763,2896,3088,3179,3384,3432,3628,3772,40
98,11066,11488,11515,11635,11949,12170,12478,12526,12645,12875,13001,13036,13295,1354
tatus predicted
F:16780,16976,17579,17602,17667,17681,17845,17899,18121,18188,18209,18336,18670,18680
F:21900,21935,22295,22495,22627,22897,23024,23318,23883,24012,24177,24290,24447,24642,
F:26171,26178,26184,26190/Binding site: phosphate (ser) (covalent) #status experiment

Query Match 96.9%; Score 31; DB 1; Length 26926;
Best Local Similarity 83.3%; Pred. No. 3.8e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEPGV 6
Db 24033 GEPGCI 24038

RESULT 8
138667
melanoma antigen Mage-8 - human
C:Species: Homo sapiens (man)
C:Date: 07-Jun-1996 #sequence, revision 07-Jun-1996 #text, change 18-Feb-2000
C:Accession: 138667
R:De Plaen, E.; Arden, K.; Traversari, C.; Gaforio, J.J.; Szikora, J.P.; De Smet, C.;
con, F.
Immunogenetics 40, 360-369, 1994
A:Title: Structure, Chromosomal localization, and expression of 12 genes of the MAGE
A:Reference number: 138659; MUID:95012457
A:Accession: 138667
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-234 <RMS>
A:Cross-references: EMBL:U10693; NID:9533525; PIDN:AAA68876.1; PID:9533526
C:Genetics:
A:Gene: GDB:MAGE8; MAGE8
A:Cross-references: GDB:331123
A:Map position: Xq28-Xq28
A:introns: #status absent
C:Superfamily: tumor associated protein MAGE

Query Match 90.6%; Score 29; DB 2; Length 234;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEPGV 6
Db 20 GEPGCI 25

RESULT 9
C82615
haloalkane denhalogenase XFI965 [Imported] - Xylella fastidiosa (strain 9a5c)

C:Species: *Xylella fastidiosa*
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: C82615
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
 A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Sep-2000
 C:Accession: F83189
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L.
 .; Lory, S.; Olson, M.V.
 A:Reference number: A82950
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: C82615
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1301 <SIM>
 A:Cross-references: GB:AE004016; GB:AE003849; NID:99107070; PIDN:AAE84767.1; GSPDB:GN001
 A:Experimental source: strain 9a5c
 R:Stimpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, H
 as-Neto, E.; Docena, C.; El-Dorry, H.; Pacicanal, A.P.; Ferreira, A.J.S.
 Submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohn
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kilaflma, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
 A:Authors: Martins, E.M.F.; Matsushima, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Savasak
 A:Authors: da Silva, A.C.R.; da Silva, E.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XFL965

Query Match 90.6%; Score 29; DB 2; Length 301;
 Best Local Similarity 83.3%; Pred. No. 1.5e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GEPGV 6
 Db 75 GDAFV 80
 RESULT 10
 S43131
 thioedoxin reductase (NADPH) (EC 1.6.4.5) - *Coxiella burnetii*
 C:Species: *Coxiella burnetii*
 C:Date: 25-Dec-1994 #sequence_revision 18-Oct-1996 #text_change 11-Jun-1999
 C:Accession: S43131
 R:Oswald, W.
 Submitted to the EMBL Data Library, November 1993
 A:Reference number: S43131
 A:Accession: S43131
 A:Molecule type: DNA
 A:Residues: 1-321 <OSM>
 A:Cross-references: EMBL:X75627; NID:9468527; PIDN:CA53288.1; PID:9468528
 C:Complex: homodimer; each chain has a tightly associated but noncovalently bound FAD
 C:Function:
 A:Description: catalyzes the reversible reduction of oxidized thioedoxin by NADPH
 C:Superfamily: thioedoxin reductase; thioedoxin reductase homology
 C:Keywords: FAD; flavoprotein; homodimer; NADP; oxidoreductase; redox-active disulfide
 F;3-315/Domain: thioedoxin reductase homology <TRAB>
 F;8-39/Region: beta-alpha-beta FAD nucleotide-binding fold
 F;149-176/Region: beta-alpha-beta NADP nucleotide-binding fold
 F;137-140/Disulfide bonds: redox-active #status predicted

Query Match 90.6%; Score 29; DB 1; Length 321;
 Best Local Similarity 83.3%; Pred. No. 1.5e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GEPGV 6
 Db 55 GEPGL 60

RESULT 11
 F83189
 Probable aminotransferase PA3659 [imported] - *Pseudomonas aeruginosa* (strain PA01)
 C:Species: *Pseudomonas aeruginosa*
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Sep-2000
 C:Accession: F83189
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L.
 .; Lory, S.; Olson, M.V.
 A:Reference number: A82950
 A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa
 A:Reference number: A82950
 A:Accession: F83189
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1402 <STO>
 A:Cross-references: GB:AE004785; GB:AE004091; NID:9949809; PIDN:AMG07047.1; GSPDB:GN
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA3659

Query Match 90.6%; Score 29; DB 2; Length 402;
 Best Local Similarity 83.3%; Pred. No. 1.9e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GEPGV 6
 Db 119 GEPGL 124

RESULT 12
 JC5178
 Probable starvation-sensing protein A - *Streptomyces coelicolor*
 C:Species: *Streptomyces coelicolor*
 C:Date: 16-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 31-Jan-2000
 C:Accession: JC5178; T42059
 R:Schneider, D.; Bruton, C.J.; Chater, K.F.
 Gene 177, 243-251, 1996
 A:Title: Characterization of spaA, a *Streptomyces coelicolor* gene homologous to a gen
 A:Reference number: JC5178; MUID:97080525
 A:Accession: JC5178
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-413 <SCCH>
 A:Cross-references: EMBL:X94190; NID:91694907; PIDN:CA63900.1; PID:e222100; PID:9169
 C:Comment: This protein is involved in an intercellular signalling system.
 C:Genetics:
 A:Gene: spaA
 A:Start codon: GTG
 F;221/Active site: Asp #status predicted
 F;247,273/Active site: Glu #status predicted

Query Match 90.6%; Score 29; DB 2; Length 413;
 Best Local Similarity 83.3%; Pred. No. 2e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GEPGV 6
 Db 376 GEPGL 381

RESULT 13
 H83347
 Hypothetical protein PA3377 [imported] - *Pseudomonas aeruginosa* (strain PA01)
 C:Species: *Pseudomonas aeruginosa*
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Sep-2000
 C:Accession: H83347
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L.
 .; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
A:Reference number: A82950
A:Accession: H83347
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-424 <STO>
A:Cross-references: GB:AE004664; GB:AE004091; NID:g9948415; PIDN:AG05765.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2377

Query Match 90.6%; Score 29; DB 2; Length 424;
Best Local Similarity 83.3%; Pred. No. 2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEPGV 6
|||||:
DB 348 GEPGL 353

RESULT 14

T35776
hypothetical protein SC8A6.09c SC8A6.09 - *Streptomyces coelicolor*
C:Species: *Streptomyces coelicolor*
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: T35776
R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1998
A:Reference number: Z21570
A:Accession: T35776
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-443 <SEE>
A:Cross-references: EMBL:AL031013; PIDN:CA19781.1; GSPDB:GN00070; SCOEDB:SC8A6.09c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC8A6.09c

Query Match 90.6%; Score 29; DB 2; Length 443;
Best Local Similarity 83.3%; Pred. No. 2.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEPGV 6
|||||:
DB 215 GEPGL 220

RESULT 15

F83268
probable PAD-dependent glycerol-3-phosphate dehydrogenase PA3025 [imported] - *Pseudomonas*
C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Sep-2000
C:Accession: F83268
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
A:Reference number: A82950
A:Accession: F83268
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-526 <STO>
A:Cross-references: GB:AE004727; GB:AE004091; NID:g9949119; PIDN:AG06413.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA3025

Query Match 90.6%; Score 29; DB 2; Length 526;

Best Local Similarity 83.3%; Pred. No. 2.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 GEPGV 6
|||||:
DB 94 GEPGL 99

Search completed: March 6, 2001, 12:50:57
Job time: 178 sec

Query Match 100.0%; Score 32; DB 1; Length 551;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEPGV 6
|||||
DB .471 GEPGV 476

RESULT 2
ABCI_HUMAN STANDARD; PRT; 2201 AA.
AC 095477; 09UN08; 09UN07; 09UN06;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ATP-BINDING CASSETTE, SUB-FAMILY A, MEMBER 1 (ATP-BINDING CASSETTE
DE TRANSPORTER 1) (ATP-BINDING CASSETTE 1) (ABC-1) (CHOLESTEROL EFFLUX
DE REGULATORY PROTEIN).
GN ABC1 OR ABC1 OR CERP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
[1]
RP SEQUENCE FROM N.A.
RA Langmann T., Klucken J., Reil M., Liebisch G., Chimini G.,
RA Kaminski W., Schmitz G.;
RT "Molecular cloning, tissue distribution and sterol regulation of
human ATP-binding cassette transporter 1 (ABCI)."
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=9364413; Pubmed=10431238;
RA Rust S., Rosier M., Funke H., Real J., Amoura Z., Piette J.-C.,
RA Deleuze J.-F., Brewer H.B., Duvergier N., Denelle P., Assmann G.;
RT "Tangier disease is caused by mutations in the gene encoding
ATP-binding cassette transporter 1.";
RL Nat. Genet. 22:352-355(1999).
[3]
RN VARIANTS TD ARG-537; LEU-633 DEL AND ARG-1417.
RX MEDLINE=9364411; Pubmed=10431236;
RA Brooks-Wilson A., Marcell M., Clee S.M., Zhang L.-H., Roomp K.,
RA van Dam M., Yu L., Brewer C., Collins J.A., Molhuizen H.O.F.,
RA Loubser O., Ouellette B.F.F., Fichter K., Ashbourne-Excoffon K.J.,
RA Samsen C.W., Scherer S., Mott S., Denis M., Martindale D.,
RA Frohlich J., Morgan K., Koop B., Plimstone S., Kastelein J.J.,
RA Hayden M.R.;
RT "Mutations in ABC1 in Tangier disease and familial high-density
lipoprotein deficiency.";
RL Nat. Genet. 22:336-345(1999).
[4]
RN VARIANTS TD SER-530; SER-875 AND VAL-877.
RX MEDLINE=9364412; Pubmed=10431237;
RA Bodzioch M., Orso E., Klucken J., Langmann T., Botcher A.,
RA Diederich W., Drobnik W., Barlage S., Buchler C., Porsch-Ozcunmez M.,
RA Kaminski W.E., Hahmann H.W., Oette K., Rothe G., Aslanidis C.,
RA Lackner K.J., Schmitz G.;
RT "The gene encoding ATP-binding cassette transporter 1 is mutated in
Tangier disease.";
RL Nat. Genet. 22:347-351(1999).
[5]
RN VARIANTS TD LEU-1229 AND HIS-1740.
RX MEDLINE=20171564; Pubmed=10706591;
RA Brousseau M.E., Schaefer E.J., Dupuis J., Eustace B.,
RA Van Bersewenh P., Goldkamp A.L., Thurston L.M., Fitzgerald M.G.,
RA Vasek-Mckenna D., O'Neill G., Eberhart G.P., Wellfischbach B.,
RA Ordoas J.M., Freeman M.W., Brown R.H. Jr., Gu J.Z.;
RT "Novel mutations in the gene encoding ATP-binding cassette 1 in four
Tangier disease kindreds.";
RL J. Lipid Res. 41:433-441(2000).
CC -1- FUNCTION: CAMP-DEPENDENT AND SULFONYLUREA-SENSITIVE ANION

CC TRANSPORTER. KEY GATEKEEPER INFLUENCING INTRACELLULAR CHOLESTEROL
CC TRANSPORT.
CC -1- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
CC ATP BINDING CASSETTE (ABC) DOMAIN.
CC -1- DISEASE: DEFECTS IN ABCA1 ARE A CAUSE OF TANGIER DISEASE (TD). TD
CC IS A RECESSIVE DISORDER CHARACTERIZED BY ABSENCE OF HIGH DENSITY
CC LIPOPROTEIN (HDL), CHOLESTEROL FROM PLASMA, HEPATOSPLENOMEGALY,
CC PERIPHERAL NEUROPATHY, AND FREQUENTLY PREMATURE CORONARY ARTERY
CC DISEASE (CHD).
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS). MDR SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ012376; CAA10005.1; -;
DR EMBL: AF165281; AAD49849.1; -;
DR EMBL: AF165306; AAD49852.1; -;
DR EMBL: AF165287; AAD49852.1; JOINED.
DR EMBL: AF165288; AAD49852.1; JOINED.
DR EMBL: AF165289; AAD49852.1; JOINED.
DR EMBL: AF165290; AAD49852.1; JOINED.
DR EMBL: AF165291; AAD49852.1; JOINED.
DR EMBL: AF165292; AAD49852.1; JOINED.
DR EMBL: AF165293; AAD49852.1; JOINED.
DR EMBL: AF165294; AAD49852.1; JOINED.
DR EMBL: AF165295; AAD49852.1; JOINED.
DR EMBL: AF165296; AAD49852.1; JOINED.
DR EMBL: AF165297; AAD49852.1; JOINED.
DR EMBL: AF165298; AAD49852.1; JOINED.
DR EMBL: AF165299; AAD49852.1; JOINED.
DR EMBL: AF165300; AAD49852.1; JOINED.
DR EMBL: AF165301; AAD49852.1; JOINED.
DR EMBL: AF165302; AAD49852.1; JOINED.
DR EMBL: AF165303; AAD49852.1; JOINED.
DR EMBL: AF165304; AAD49852.1; JOINED.
DR EMBL: AF165305; AAD49852.1; JOINED.
DR EMBL: AF165309; AAD49854.1; -;
DR EMBL: AF165307; AAD49854.1; JOINED.
DR EMBL: AF165308; AAD49854.1; JOINED.
DR EMBL: AF165310; AAD49853.1; -;
DR MIM: 600046; -;
DR MIM: 205400; -;
DR INTERPRO: IPR001617; -;
DR PFAM: PF00005; ABC_tran; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Glycoprotein; Transmembrane; Transport;
KW Disease mutation.
KW FT TRANSSEM 580 596 POTENTIAL.
FT TRANSSEM 630 646 POTENTIAL.
FT TRANSSEM 657 673 POTENTIAL.
FT TRANSSEM 689 705 POTENTIAL.
FT TRANSSEM 711 727 POTENTIAL.
FT TRANSSEM 981 997 POTENTIAL.
FT TRANSSEM 1291 1307 POTENTIAL.
FT TRANSSEM 1601 1617 POTENTIAL.
FT TRANSSEM 1648 1664 POTENTIAL.
FT TRANSSEM 1677 1693 POTENTIAL.
FT TRANSSEM 1715 1731 POTENTIAL.
FT TRANSSEM 1794 1810 POTENTIAL.
FT NP_BIND 873 880 ATP (POTENTIAL).
FT NP_BIND 1886 1893 ATP (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 101 101 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

FT CARBOHYD 232 232 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 429 429 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 461 461 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 760 760 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1084 1084 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1234 1234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1393 1393 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1444 1444 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1577 1577 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1984 1984 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2178 2178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 339 339 V -> A (IN TD).
FT VARIANT 527 527 R -> W (IN TD).
FT VARIANT 530 530 W -> S (IN TD).
FT VARIANT 537 537 O -> R (IN TD).
FT VARIANT 537 537 MISSING (IN TD).
FT VARIANT 633 633 MISSING (IN TD).
FT VARIANT 633 633 MISSING (IN TD).
FT VARIANT 875 875 N -> S (IN TD).
FT VARIANT 877 877 A -> V (IN TD).
FT VARIANT 877 877 /FTID=VAR_009150.
FT VARIANT 1229 1229 /FTID=VAR_009151.
FT VARIANT 1417 1417 D -> L (IN TD).
FT VARIANT 1417 1417 /FTID=VAR_009152.
FT VARIANT 1457 1457 C -> R (IN TD).
FT VARIANT 1457 1457 /FTID=VAR_009153.
FT VARIANT 1740 1740 I -> R (IN TD).
FT VARIANT 1740 1740 /FTID=VAR_009154.
FT CONFLICT 1495 1495 N -> H (IN TD).
FT CONFLICT 1527 1527 /FTID=VAR_009155.
FT CONFLICT 1588 1588 T -> I (IN AAD49852).
FT CONFLICT 1685 1685 R -> K (IN AAD49852).
FT CONFLICT 2108 2108 P -> L (IN AAD49852).
FT CONFLICT 2201 2201 P -> L (IN AAD49853).
SO SEQUENCE 2201 AA; 246987 MW; 1EC5081808AFB520 CRC64;

Query Match 100.0%; Score 32; DB 1; Length 2201;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEAPGV 6
DB 28 GEAPGV 33

RESULT 3
ABCI_MOUSE STANDARD: PRT: 2201 AA.
AC P41233:
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ATP-BINDING CASSETTE, SUB-FAMILY A, MEMBER 1 (ATP-BINDING CASSETTE
DE TRANSPORTER 1) (ATP-BINDING CASSETTE 1) (ABC-1).
GN ABCA1 OR ABC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DA/2; TISSUE=MACROPHAGE;
RX MEDLINE=94375008; Pubmed=8086782;
RA Luciani M.F., Denizot F., Savary S., Mattei M.-G., Chimi G.;
"Cloning of two novel ABC transporters mapping on human chromosome

```

```

RT 9."
RL Genomics 21:150-159(1994).
CC -1- FUNCTION: CAMP-DEPENDENT AND SULFONYLUREA-SENSITIVE ANION
CC TRANSPORTER. KEY GATEKEEPER INFLUENCING INTRACELLULAR CHOLESTEROL
CC TRANSPORT (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED IN ADULT TISSUES. HIGHEST
CC LEVELS ARE FOUND IN PREGNANT UTERUS AND UTERUS.
CC -1- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
CC ATP BINDING CASSETTE (ABC) DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS). MDR SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: X75926; CAA53530.1;
CC MGD: MGI:99607; ABCA1.
DR INTERPRO: IPR001617;
DR PFM: PFM00005; ABC_tran; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Glycoprotein; Transmembrane; Transport.
FT TRANSMEM 580 596 POTENTIAL.
FT TRANSMEM 630 646 POTENTIAL.
FT TRANSMEM 657 673 POTENTIAL.
FT TRANSMEM 689 705 POTENTIAL.
FT TRANSMEM 711 727 POTENTIAL.
FT TRANSMEM 981 997 POTENTIAL.
FT TRANSMEM 1291 1307 POTENTIAL.
FT TRANSMEM 1601 1617 POTENTIAL.
FT TRANSMEM 1648 1664 POTENTIAL.
FT TRANSMEM 1677 1693 POTENTIAL.
FT TRANSMEM 1715 1731 POTENTIAL.
FT TRANSMEM 1794 1810 POTENTIAL.
FT NP BIND 873 880 ATP (POTENTIAL).
FT NP BIND 1886 1893 ATP (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 101 101 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 232 232 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 429 429 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 461 461 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 760 760 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1084 1084 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1234 1234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1393 1393 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1439 1439 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1444 1444 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1577 1577 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1984 1984 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2178 2178 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 2201 AA; 246686 MW; C7DB9A6CB7A5DB73 CRC64;

```

```

Query Match 100.0%; Score 32; DB 1; Length 2201;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 GEAPGV 6
DB 28 GEAPGV 33

```

RESULT 4
GSPH_KLEPN STANDARD: PRT: 170 AA.
ID GSPH_KLEPN
AC P15747;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE GENERAL SECRETION PATHWAY PROTEIN H PRECURSOR (PULLULANASE SECRETION
DE PROTEIN PULH).
GN PULH
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN UNF 5023;
RX MEDLINE=91109698; PubMed=2129543;
RA Reys I., Pugsley A.P.;
RT "Five additional genes in the pulC-O operon of the gram-negative
RT bacterium Klebsiella oxytoca UNF5023 which are required for
RT pullulanase secretion."
RL Mol. Gen. Genet. 222:176-184(1990).
CC -1- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
CC EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLOCATION OF PULLULANASE.
CC -1- SIMILARITY: BELONGS TO THE PULH/OUTH/XPSH/EXEH/XCPU FAMILY.
CC -----
CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M32613; AAA25130.1; ALT_INIT.
DR PIR: S11918; S11918.
DR INTERPRO: IPR001120;
DR INTERPRO: IPR002416;
DR PRINTS: PRO00885; BCTERIALGSPH.
DR PROSITE: PS00409; PROKAR_NTER_METHYL; 1.
KW Transport; Methylation.
FT PROPEP 1 5 BY SIMILARITY.
FT CHAIN 6 170 GENERAL SECRETION PATHWAY PROTEIN H.
FT MOD_RES 6 6 METHYLATION (BY SIMILARITY).
SQ SEQUENCE 170 AA; 18403 MW; 020CC4D0636280 CRC64;

Query Match 96.98; Score 31; DB 1; Length 170;
Best Local Similarity 83.3%; Pred. NO. 15;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GARGV 6
DB 149 GARGV 154

RESULT 5
PGCA_BOVIN STANDARD: PRT: 2364 AA.
ID PGCA_BOVIN
AC P13608; 028159; P79117;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE AGGREGAN CORE PROTEIN PRECURSOR (CARTILAGE-SPECIFIC PROTEOGLYCAN CORE
DE PROTEIN) (CSPCP).
GN AGC1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A.

RA Hering T.M., Kollar J., Huynh T.D.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE OF 563-1056 FROM N.A.
RX MEDLINE=89380219; PubMed=2528543;
RA Antonsson P., Heinegaard D., Oldberg A.;
RT "The keratan sulfate-enriched region of bovine cartilage proteoglycan
RT consists of a consecutively repeated hexapeptide motif."
RL J. Biol. Chem. 264:16170-16173(1989).
RN [3]
RN SEQUENCE OF 1609-2113 AND 2151-2364 FROM N.A.
RX MEDLINE=87270630; PubMed=3111460;
RA Oldberg A., Antonsson P., Heinegaard D.;
RT "The partial amino acid sequence of bovine cartilage proteoglycan,
RT deduced from a cDNA clone, contains numerous Ser-Gly sequences
RT arranged in homologous repeats."
RL Biochem. J. 243:255-259(1987).
RN [4]
RN SEQUENCE OF 2114-2150 FROM N.A.
RC TISSUE=CARTILAGE;
RX MEDLINE=93352525; PubMed=8349621;
RA Fuelleop C., Walcz E., Valyon M., Glant T.T.;
RT "Expression of alternatively spliced epidermal growth factor-like
RT domains in aggregates of different species. Evidence for a novel
RT module."
RL J. Biol. Chem. 268:17377-17383(1993).
RN [5]
RN PARTIAL SEQUENCE.
RX MEDLINE=85027710; PubMed=6489519;
RA Perin J.P., Bonnet F., Jolles J., Jolles P.;
RT "Sequence data concerning the protein core of the cartilage
RT proteoglycan monomers. Characterization of a sequence allowing the
RT synthesis of an oligonucleotide probe."
RL FEBS Lett. 176:37-42(1984).
RN [6]
RN PARTIAL SEQUENCE.
RX MEDLINE=87005253; PubMed=3530809;
RA Perin J.P., Bonnet F., Jolles P.;
RT "Structural relationship between link proteins and proteoglycan
RT monomers."
RL FEBS Lett. 206:73-77(1986).
CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
CC MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
CC IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A
CC REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
CC -1- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
CC TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
CC MAKES UP THE C-TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
CC CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
CC THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
CC AND G3.
CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO N-LINKED AND
CC O-LINKED (ABOUT 40) OLIGOSACCHARIDES.
CC -1- PTM: THE KERATAN SULFATE CONTENTS DIFFER CONSIDERABLY BETWEEN
CC ADULT AND FETAL BOVINE PROTEOGLYCAN.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 4 LINK DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEAT.
CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

```

CC EMBL: U76615; AAB38524.1; -
DR EMBL: L07053; -; NOT_ANNOTATED_CDS.
DR PIR: A27752; A27752.
DR PIR: A29164; A29164.
DR PIR: B29164; B29164.
DR PIR: E29164; E29164.
DR PIR: G27751; G27751.
DR HSSP: P00740; 11XA.
DR INTERPRO: IPR000152; -
DR INTERPRO: IPR000436; -
DR INTERPRO: IPR000538; -
DR INTERPRO: IPR000561; -
DR INTERPRO: IPR001304; -
DR INTERPRO: IPR001881; -
DR PFAM: PF00008; EGF_1.
DR PFAM: PF00193; Xlink_4.
DR PFAM: PF00059; lectin_C_1.
DR PFAM: PF00084; sushi_1.
DR PROSITE: PS00010; ASX_HYDROXYL_1.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS01187; EGF_CA_1.
DR PROSITE: PS01241; LINK_4.
DR PROSITE: PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE: PS50041; C-TYPE_LECTIN_2; 1.
KW Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
EGF-like domain; Alternative splicing; Repeat; Immunoglobulin domain.
FT SIGNAL 1 16
FT CHAIN 17 2364
FT DOMAIN 44 140
FT DOMAIN 170 247
FT DOMAIN 268 349
FT DOMAIN 504 581
FT DOMAIN 602 683
FT DOMAIN 774 907
FT DOMAIN 1433 2112
FT DOMAIN 2113 2239
FT DOMAIN 2149 2364
FT DOMAIN 2114 2364
FT DOMAIN 2150 2276
FT REPEAT 2280 2338
FT CARBOHYD 126 126
FT CARBOHYD 239 239
FT CARBOHYD 333 333
FT CARBOHYD 387 387
FT CARBOHYD 611 611
FT CARBOHYD 667 667
FT DISULFID 51 133
FT DISULFID 175 246
FT DISULFID 199 220
FT DISULFID 273 348
FT DISULFID 297 318
FT DISULFID 509 580
FT DISULFID 533 554
FT DISULFID 607 682
FT DISULFID 631 652
FT DISULFID 2117 2128
FT DISULFID 2182 2274
FT DISULFID 2250 2266
FT DISULFID 2281 2324
FT DISULFID 2310 2337
FT VASPLIC 2114 2150
SQ SEQUENCE 2364 AA: 246359 MW: 66F83763420C3D4C CRC64;

```

```

RESULT 6
MAG8_HUMAN STANDARD: PRT: 234 AA.
ID MAG8_HUMAN
AC P43561.
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE MELANOMA-ASSOCIATED ANTIGEN 8 (MAGE-8 ANTIGEN).
GN MAGE8 OR MAGE8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95012457; PubMed=7927540;
RA de Plaen E., Arden K., Traversari C., Gaforio J.J., Sikora J.-P.,
RA Brasseur R., Chomez P., van der Bruggen P., Lethé B., Lurquin C.,
RA "Structure, chromosomal localization, and expression of 12 genes of
RT the MAGE family";
RL Immunogenetics 40:360-369(1994).
CC -!- FUNCTION: NOT KNOWN, THOUGH MAY PLAY A ROLE IN EMBRYONAL
CC DEVELOPMENT AND TUMOR TRANSFORMATION OR ASPECTS OF TUMOR
CC PROGRESSION.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN MANY TUMORS OF SEVERAL TYPES,
CC SUCH AS MELANOMA, HEAD AND NECK SQUAMOUS CELL CARCINOMA, LUNG,
CC CARCINOMA AND BREAST CARCINOMA, BUT NOT IN NORMAL TISSUES EXCEPT
CC FOR TESTES AND PLACENTA.
CC -!- SIMILARITY: BELONGS TO THE MAGE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U10693; AAB6876.1; -
DR INTERPRO: IPR002190; -
DR PFAM: PF01454; MAGE_1.
KW Antigen; Multigene family; Tumor antigen.
FT DOMAIN 40 43
FT SEQUENCE 234 AA: 25197 MW: 058A92EE6003A982 CRC64;

```

```

Query Match 96.9%; Score 31; DB 1; Length 2364;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 GEPGV 6
DB 1624 GEPGL 1629

```

```

Query Match 90.6%; Score 29; DB 1; Length 234;
Best Local Similarity 83.3%; Pred. No. 52;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 GEPGV 6
DB 20 GEPGL 25

```

```

RESULT 7
TRXB_COXBU STANDARD: PRT: 321 AA.
ID TRXB_COXBU
AC P39916.
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE THIODEDOLIN REDUCTASE (EC 1.6.4.5) (TRXB).
GN TRXB.
OS Coxiella burnetii.
OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;
OC Coxiella group; Coxiella.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NINE MILE PHASE I / BRATISLAVA;
RA Oswald W.;

```

```

RL Thesis (1994), Justus Liebig University / Frankfurt, Germany.
CC -1- CATALYTIC ACTIVITY: NADPH + OXIDIZED THIOREDOXIN = NADP(+) +
CC REDUCED THIOREDOXIN.
CC -1- COFACTOR: FAD.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE
CC OXIDOREDUCTASES CLASS-II.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X75627; CA53288.1; -.
DR HSSP: P09625; 1TDF.
DR INTERPRO: IPR00103; -.
DR PRINTS: PR00469; PYRIDINE_RED0X.2; 1.
DR PROSITE: PS00573; PYRIDINE_RED0X.2; 1.
KW Redox-active center; Oxidoreductase; NADP; Flavoprotein; FAD.
FT NP_BIND 7 22 FAD (ADP PART) (PROBABILE).
FT DISULFID 137 140 REDOX-ACTIVE (BY SIMILARITY).
FT NP_BIND 149 163 NAD(P) (BY SIMILARITY).
FT NP_BIND 278 288 FAD (FLAVIN PART) (BY SIMILARITY).
SQ SEQUENCE 321 AA; 34660 MW; 7F9CA07B282BC0E5 CRC64;

```

```

Query Match          90.6%; Score 29; DB 1; Length 321;
Best Local Similarity 83.3%; Pred. No. 70;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GEARGV 6
   11111
DB 55 GEARGL 60

```

```

RESULT 8
NR42_XENLA STANDARD; PRT; 577 AA.
AC 004913;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ORPHAN NUCLEAR RECEPTOR NURR1 (NERVE GROWTH FACTOR INDUCED PROTEIN T-B
DE HOMOLOG).
GN NR42.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Plagidae;
OC Xenopodinae; Xenopus.
CC [1]
CC SEQUENCE FROM N.A.
CC TISSUE=NEURULA;
CC MEDLINE=93277961; Pubmed=8504173;
CC Smith T.S., Matharu P.J., Sweeney G.E.;
CC "Cloning and sequencing of a Xenopus homologue of the inducible
CC orphan receptor NgR1-B.";
CC Biochem. Biophys. Acta 1173:239-242(1993).
CC -1- FUNCTION: PROBABLE NUCLEAR RECEPTOR.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR4 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X70700; CA50031.1; -.
DR PIR: S33763; S33763.
DR PIR: S33713; S31713.
DR HSSP: P19793; 2NML.
DR INTERPRO: IPR000536; -.
DR INTERPRO: IPR001628; -.
DR INTERPRO: IPR003070; -.
DR PFAM: PF00104; hormone_rec.1.
DR PFAM: PF00105; zf-C4; 1.
DR PRINTS: PR00047; STRODIFINGER.
DR PRINTS: PR01284; NUCLEARREPT.
DR PROSITE: PS00031; NUCLEAR_RECEPTOR.1.
KW DNA-binding; Transcription regulation; Nuclear protein; Receptor;
KW Zinc-finger.
FT DNM_BIND 246 311 C4-TYPE ZINC FINGERS (TWO).
FT ZN_FING 246 266 C4-TYPE.
FT ZN_FING 282 311 C4-TYPE.
FT DOMAIN 354 404 LIGAND-BINDING (BY POTENTIAL).
SQ SEQUENCE 577 AA; 64383 MW; 3A6B26EA20E8F81 CRC64;

```

```

Query Match          90.6%; Score 29; DB 1; Length 577;
Best Local Similarity 83.3%; Pred. No. 1,2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GEARGV 6
   11111
DB 215 GEARPV 220

```

```

RESULT 9
YAGR_ECOLI STANDARD; PRT; 732 AA.
AC P77489;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 78.1 KDA PROTEIN IN INTF-EAHI INTERGENIC REGION.
GN YAGR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=K12 / MG1655;
CC MEDLINE=9742617; Pubmed=9278503;
CC Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
CC Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
CC Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
CC Mau B., Shao Y.;
CC "The complete genome sequence of Escherichia coli K-12.";
CC Science 277:1453-1474(1997).
CC [2]
CC SEQUENCE FROM N.A.
CC Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,
CC Federfeld N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,
CC Nemeth A., Oefner P., Roberts D., Schramm S., Davis R.W.;
CC Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: YAGR, YAGS AND YAGT COULD BE THREE SUBUNITS OF A
CC DEHYDROGENASE.
CC -1- SIMILARITY: TO VERTEBRATE XANTHINE DEHYDROGENASES AND TO
CC R.PALUSIRIS 4-HYDROXYBENZOYL-COA REDUCTASE SUBUNIT HBAC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

DR EMBL; AE000136; AAC73387.1; -
 DR EMBL; U73857; AAB18013.1; -
 DR ECGENE; EG13557; YAGR.
 DR INTERPRO; IPR000674; -
 DR PFAM; PF01315; Ald_Xan_dh.C; 1.
 KW Hypothetical protein, Oxidoreductase
 SQ SEQUENCE 732 AA; 78088 MW; AF198715794F0138 CRC64;

Query Match 90.6%; Score 29; DB 1; Length 732;
 Best Local Similarity 83.3%; Pred. No. 1.5e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEPGV 6
 |||||
 Db 353 GEPGL 358

RESULT 10
 ABL2_HUMAN STANDARD; PRT; 1182 AA.
 ID ABL2_HUMAN
 AC P42684;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE TYROSINE-PROTEIN KINASE ABL2 (EC 2.7.1.112) (TYROSINE KINASE ARG).
 GN ABL2 OR ARG OR ABL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90332670; PubMed=2198571;
 RA Krub G.D., Perego R., Miki T., Aaronson S.A.;
 RT "The complete coding sequence of arg defines the Ablastin subfamily of
 cytoplasmic tyrosine kinases.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:5802-5806(1990).
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
 PROTEIN TYROSINE PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1A AND 1B (SHOWN HERE); ARE
 PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED.
 CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
 DOMAIN. BELONGS TO THE ABL SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M35296; AAA35553.1; -
 DR HSSP; P00520; IABO.
 DR MIR; 164690; -
 DR INTERPRO; IPR000719; -
 DR INTERPRO; IPR000980; -
 DR INTERPRO; IPR001245; -
 DR INTERPRO; IPR001452; -
 DR PFAM; PF00017; SH2; 1.
 DR PFAM; PF00018; SH3; 1.
 DR PFAM; PF00069; PKINASE; 1.
 DR PRINTS; PR00109; TYRKINASE.
 DR PRINTS; PR00401; SH2DOMAIN.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS50011; SH2; 1.
 DR PROSITE; PS50002; SH3; 1.

KW Transferase; Tyrosine-protein kinase; Proto-oncogene; ATP-binding;
 KW Phosphorylation; SH2 domain; SH3 domain; Alternative splicing.
 FT DOMAIN 107 167
 FT DOMAIN 173 263
 FT DOMAIN 288 539
 FT DOMAIN 561 564
 FT DOMAIN 658 660
 FT DOMAIN 732 739
 FT DOMAIN 843 1055
 FT DOMAIN 984 988
 FT NP_BIND 294 302
 FT BINDING 317 317
 FT ACT_SITE 409 409
 FT MOD_RES 439 439
 FT VARSPPLIC 1 73

SEQUENCE 1182 AA; 128343 MW; ED93869BC2B14FPA CRC64;

Query Match 90.6%; Score 29; DB 1; Length 1182;
 Best Local Similarity 83.3%; Pred. No. 2.4e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEPGV 6
 |||||
 Db 9 GEPGL 14

RESULT 11
 PCGA_CHICK STANDARD; PRT; 2109 AA.
 ID PCGA_CHICK
 AC P07898; Q90991; Q90820; Q91047; Q90810;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE AGGRECAN CORE PROTEIN PRECURSOR (CARTILAGE-SPECIFIC PROTEOGLYCAN CORE
 DE PROTEIN) (CSPCP).
 GN AGC1.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WHITE LEGHORN; TISSUE=EMBRYO;
 RX MEDLINE=94043149; PubMed=8226878;
 RA Li H., Schwartz N.B., Vertel B.M.;
 RT "cDNA cloning of chick cartilage chondroitin sulfate (aggrecan) core
 RT protein and identification of a stop codon in the aggrecan gene
 RT associated with the chondrodys trophy, nanomelia.";
 RL J. Biol. Chem. 268:23504-23511(1993).
 RN [2]
 RP SEQUENCE OF 1042-1559 FROM N.A.
 RC TISSUE=EMBRYO;
 RX MEDLINE=90307744; PubMed=1694853;
 RA Krueger R.C., Jr., Fields T.A., Mensch J.R., Jr., Schwartz N.B.;
 RT "Chick cartilage chondroitin sulfate proteoglycan core protein. II.
 RT Nucleotide sequence of cDNA clone and localization of the S103L
 RT epitope.";
 RL J. Biol. Chem. 265:12088-12097(1990).
 RN [3]
 RP SEQUENCE OF 1-1855 AND 1893-2109 FROM N.A.
 RC TISSUE=CARTILAGE;
 RX MEDLINE=93111968; PubMed=1339285;
 RA Chandrasekaran L., Tanzer M.L.;
 RT "Molecular cloning of chicken aggrecan. Structural analyses.";
 RL Biochem. J. 268:903-910(1992).
 RN [4]
 RP ERRATUM.
 RX MEDLINE=94107258; PubMed=8280087;
 RA Chandrasekaran L., Tanzer M.L.;


```

FT CONFLICT 1988 1988 F -> S (IN REF. 6).
SO SEQUENCE 2109 AA; 223492 MW; 7F824FD5B3A2ABDA CRC64;

Query Match
Best Local Similarity 90.6%; Score 29; DB 1; Length 2109;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEPAG 6
   1:|||||
Db 687 GDAFV 692

RESULT 12
PRL_PROAT STANDARD; PRT: 200 AA.
ID PRL_PROAT
AC P33091;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PROLACTIN (PRL).
OS Protopterus aethiopicus (Marbled lungfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Dipnoi; Lepidosireniformes; Protopteridae; Protopterus.
RN [1]
RP SEQUENCE.
RX MEDLINE=93320105; PubMed=6329446;
RA Noso T., Nicoll C.S., Kawachi H.;
RT "Lungfish prolactin exhibits close tetrapod relationships.";
RL Biochim. Biophys. Acta 1164:159-165(1993).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: PITUITARY GLANDS.
CC -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
DR HSSP: 028632; IAN3.
DR INTERPRO: IPR001400; -.
DR PFAM: PF00103; hormone; 1.
DR PRINTS: PR00836; SOMATOTROPIN.
DR PROSITE: PS00266; SOMATOTROPIN_1; 1.
DR PROSITE: PS00338; SOMATOTROPIN_2; 1.
KW Hormone; Pituitary.
FT DISULFID 4 11 BY SIMILARITY.
FT DISULFID 59 175 BY SIMILARITY.
FT DISULFID 192 200 BY SIMILARITY.
FT VARIANT 22 22 R -> F.
FT VARIANT 41 41 F -> G.
SQ SEQUENCE 200 AA; 22904 MW; BEBA3E9B300443B CRC64;

Query Match
Best Local Similarity 87.5%; Score 28; DB 1; Length 200;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEPAG 5
   1:|||||
Db 108 GEPAG 112

RESULT 13
BPHF_RHOSO STANDARD; PRT: 258 AA.
ID BPHF_RHOSO
AC 005151;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE 4-HYDROXY-2-OXOVALERATE ALDOOLASE (EC 4.1.1.2.-).
GN BPHF OR ETBF.
OS Rhodococcus sp. (strain RHA1).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=97225808; PubMed=9073078;
RA Masai E., Sugiyama K., Iwashita N., Shimizu S., Hauschild J.E.,
RA Hata T., Kimbara K., Yano K., Fukuda M.;
RT "The bphf meta-cleavage pathway genes involved in
RT biphenyl/polychlorinated biphenyl degradation are located on a linear
RT plasmid and separated from the initial bphb genes in Rhodococcus sp.
RT strain RHA1.";
RL Gene 187:141-149(1997).
CC -1- FUNCTION: CATALYZES THE PRODUCTION OF PYRUVATE AND ACETALDEHYDE
CC FROM 4-HYDROXY-2-OXOVALERATE.
CC -1- PATHWAY: SIXTH STEP IN THE META-CLEAVAGE OF POLYCHLOROBIPHENYLS
CC (PCB) PATHWAY.
CC -1- INDUCTION: BY GROWTH ON ETHYLENE OR BIPIHENYL.
CC -1- SIMILARITY: BELONGS TO THE HPCH/HPAI ALDOOLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D78322; BAA18937.1; -.
DR Aromatic hydrocarbons catabolism; Lyase; Plasmid.
KW SEQUENCE 258 AA; 27159 MW; A539C46C4DA0190F CRC64;
SQ SEQUENCE 258 AA; 27159 MW; A539C46C4DA0190F CRC64;

```

```

Query Match
Best Local Similarity 87.5%; Score 28; DB 1; Length 258;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEPAG 6
   1:|||||
Db 208 GKAPV 213

```

```

RESULT 14
YT32_STRFR STANDARD; PRT: 305 AA.
ID YT32_STRFR
AC P20185;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 32.6 KDA PROTEIN IN TRANSPOSON TN4556.
OS Streptomyces fradiae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RP TRANSPOSON-TN4556;
RX MEDLINE=90185236; PubMed=2155856;
RA Slemieniak D.R., Slightom J.L., Chung S.T.;
RT "Nucleotide sequence of Streptomyces fradiae transposable element
RT Tn4556: a class-II transposon related to Tn3.";
RL Gene 86:1-9(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M29297; AAA88566.1; -.
DR PIR: J00428; J00428.
KW Hypothetical protein; Transposable element.
SQ SEQUENCE 305 AA; 32613 MW; 20C1965D9A84D825 CRC64;

```

```

Query Match
Best Local Similarity 87.5%; Score 28; DB 1; Length 305;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GEARPG 5
 Db 283 GEARPG 287

Db 294 GEARPG 299

Search completed: March 6, 2001, 12:54:41
 Job time: 398 sec

RESULT 15
 ID AMY1_HORVU STANDARD; PRT; 438 AA.
 AC P00693;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ALPHA-AMYLASE TYPE A ISOZYME PRECURSOR (EC 3.2.1.1) (1,4-ALPHA-D-
 DE GLUCAN GLUCANOHYDROLASE) (AMY1) (LOW PI ALPHA-AMYLASE).
 GN AMY1.1.
 OS Hordeum vulgare (Barley).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. HIMALAYA;
 RX MEDLINE=83238423; PubMed=6190808;
 RA Rogers J.C., Millman C.;
 RT Isolation and sequence analysis of a barley alpha-amylase cDNA
 RT clone.";
 RL J. Biol. Chem. 258:8169-8174(1983).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-ALPHA-GLUCOSIDIC
 CC LINKAGES IN OLIGOSACCHARIDES AND POLYSACCHARIDES.
 CC -1- COFACTOR: BINDS A CALCIUM ION REQUIRED FOR ITS ACTIVITY.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1- DEVELOPMENTAL STAGE: PRODUCTION OF ALPHA-AMYLASE IS HORMONALLY
 CC REGULATED. GERMINATING EMBRYOS PRODUCE THE HORMONE GIBBERELLIC
 CC ACID, WHICH WITHIN 10 HOURS STIMULATES THE ALEURONE CELLS COVERING
 CC THE ENDOSPERM OF THE SEED TO PRODUCE ALPHA-AMYLASE. THE ENZYME
 CC THEN DEGRADES THE STARCH WITHIN THE ENDOSPERM FOR USE BY THE
 CC DEVELOPING PLANT EMBRYO.
 CC -1- MISCELLANEOUS: THERE ARE AT LEAST 4 TYPES OF ALPHA-AMYLASE IN
 CC BARLEY.
 CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: J01236; AAA32929.1; -
 CC DR PIR: A00846; ALBH.
 CC DR HSSP: P04063; LAMY.
 CC DR MENDEL, 8617; HORVU; Amy1.1.
 CC DR INTERPRO: IPR000461; -
 CC DR PFAM: PF00128; alpha-amylase; 1.
 CC DR PRINTS: PRO0110; ALPHAAMYLASE.
 CC KW Hydrolase; Glycosidase; Carbohydrate metabolism; Seed; Germination;
 CC Calcium; Multigene family; Signal.
 CC FT SIGNAL 1 24
 CC FT CHAIN 25 438
 CC FT ACT_SITE 204 204 BY SIMILARITY.
 CC FT ACT_SITE 229 229 BY SIMILARITY.
 CC FT ACT_SITE 315 315 BY SIMILARITY.
 CC SQ SEQUENCE 438 AA; 47796 MF; 2393FDAC51E80F51 CRC64;

Query Match 87.5%; Score 28; DB 1; Length 438;
 Best Local Similarity 83.3%; Pred. No. 1.5e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEARPG 6

THIS PAGE BLANK (USPTO)


```

DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE IMMOBILIZATION ANTIGEN PRECURSOR (FRAGMENT).
OS Ichthyophthirius multifiliis.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Ophryoglenina; Ichthyophthirius.
OX NCBI_TaxID=5932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GEORGIA;
RX MEDLINE=9235298; PubMed=1631132;
RA Clark T.G., McGraw R.A., Dickerson H.W.;
RT "Developmental expression of surface antigen genes in the parasitic
RL ciliate Ichthyophthirius multifiliis."
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GEORGIA;
RX MEDLINE=93020590; PubMed=1383510;
RA Lin T.L., Dickerson H.W.;
RT "Purification and partial characterization of immobilization antigens
RL from Ichthyophthirius multifiliis."
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=GEORGIA;
RA Clark T.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=GEORGIA;
RA Clark T.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; M92907; AAC36158.1; -
KW Signal.
FT SIGNAL.
SQ SEQUENCE 394 AA; 39495 MW; 3013C2B2BEFDB682 CRC64;

Query Match
Best Local Similarity 100.0%; Score 32; DB 5; Length 394;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEPGV 6
Db 123 GEPGV 128

RESULT 3
Q9XGZ2 PRELIMINARY; PRT; 442 AA.
AC Q9XGZ2;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE IMMOBILIZATION ANTIGEN PRECURSOR.
OS IAG48.
OS Ichthyophthirius multifiliis.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Ophryoglenina; Ichthyophthirius.
OX NCBI_TaxID=5932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GI;
RX MEDLINE=99196987; PubMed=10095108;
RA Clark T.G., Lin T.L., Jackwood D.A., Sherrill J., Lin Y.,
RA Dickerson H.W.;
RT "The gene for an abundant parasite coat protein predicts tandemly
RL repetitive metal binding domains."
RN [2]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=GI;
RA Gaertig J., Gao Y., Tishgarten T., Clark T.G., Dickerson H.W.;
RT "Surface display of a parasite antigen in the ciliate Tetrahymena
RL thermophila."
RN [1]
RP Net. Biotechnol. 0:0-0(1999).
DR EMBL; AF140273; AAD31283.1; -
KW Signal.
FT SIGNAL.
SQ SEQUENCE 442 AA; 45025 MW; 52658F3F65D27AFA CRC64;

Query Match
Best Local Similarity 100.0%; Score 32; DB 5; Length 442;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEPGV 6
Db 143 GEPGV 148

RESULT 4
Q45340 PRELIMINARY; PRT; 1010 AA.
AC Q45340;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE BRKA.
GN BRKA.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TOHAMA (BP338);
RX MEDLINE=95012680; PubMed=7927748;
RA Fernandez R.C., Weiss A.A.;
RT "Cloning and sequencing of a Bordetella pertussis serum resistance
RL locus."
DR EMBL; U12276; AAA51646.1; -
SQ SEQUENCE 1010 AA; 103377 MW; 608A006EC3087B52 CRC64;

Query Match
Best Local Similarity 100.0%; Score 32; DB 2; Length 1010;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEPGV 6
Db 185 GEPGV 190

RESULT 5
Q9W6E1 PRELIMINARY; PRT; 1290 AA.
AC Q9W6E1;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE NEUROCAN CORE PROTEIN PRECURSOR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OX Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RA Li H., Leung T., Balsamo J., Hoffman S., Lillen J.;
RT "cDNA cloning of chicken neurocan and its role in regulating N-

```

RT cadherin function in embryonic chicken retina."
 RL Submitted (May-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF116836; AAD24546.2; -.
 DR HSSP; P08709; 1BF9.
 DR INTERPRO: IPR000152; -.
 DR INTERPRO: IPR000436; -.
 DR INTERPRO: IPR000538; -.
 DR INTERPRO: IPR000561; -.
 DR INTERPRO: IPR000742; -.
 DR INTERPRO: IPR001304; -.
 DR INTERPRO: IPR001438; -.
 DR INTERPRO: IPR001881; -.
 DR PFAM; PF00008; EGF_2.
 DR PFAM; PF00059; lectin_c; 1.
 DR PFAM; PF00084; sushi; 1.
 DR PFAM; PF00193; Xlink; 2.
 DR PRINTS; PR00010; EGFBL00.
 DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
 DR PROSITE; PS00615; C-TYPE-LECTIN_1; 1.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS01241; LINK; 2.
 DR PROSITE; PS50041; C-TYPE-LECTIN_2; 1.
 KW Glycoprotein; EGF-like domain.
 SO SEQUENCE 1290 AA; 138876 MW; 182BD86D0E40BE78 CRC64;

Query Match 100.0%; Score 32; DB 13; Length 1290;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEPGV 6
 DB 327 GEPGV 332

RESULT 6
 O9NOV4 PRELIMINARY; PRT; 2261 AA.
 AC O9NOV4.
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE ABCA1.
 GN ABCA1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Santamarina-Fojo S., Peterson K.M., Knapper C.L., Qiu Y.,
 RA Freeman L.A., Cheng J.-F., Osorio J., Remaley A.T., Yang X.-P.,
 RA Haudenschild C.C., Prades C., Chhimin G., Blackmon E.B.,
 RA Francos T.L., Diverger N., Rubin E.M., Rosier M., Denefle P.,
 RA Fredrickson D.S., Brewer H.B. Jr.;
 RT "Complete genomic sequence of the human ABCA1 gene: Analysis of the
 human and mouse ATP-binding cassette A promoter."
 RL Proc. Natl. Acad. Sci. U.S.A. 97:7987-7992(2000).
 DR EMBL; AF275948; AAF86276.1; -.
 SO SEQUENCE 2261 AA; 254324 MW; BA27D9B217ACAA33 CRC64;

Query Match 100.0%; Score 32; DB 4; Length 2261;
 Best Local Similarity 100.0%; Pred. No. 5.3e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEPGV 6
 DB 88 GEPGV 93

RESULT 7
 ID 062623 PRELIMINARY; PRT; 719 AA.
 AC 062623;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE CARTILAGE PROTEOGLYCAN (FRAGMENT).
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-CARTILAGE;
 RA Antonsson P.;
 RL Submitted (SEP-1988) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-CARTILAGE;
 RX MEDLINE-87270630; PubMed-3111460;
 RA Oldberg A., Antonsson P., Heinegard D.;
 RT "The partial amino acid sequence of bovine cartilage proteoglycan,
 deduced from a cDNA clone, contains numerous Ser-Gly sequences
 arranged in homologous repeats."
 RL Biochem. J. 243:255-259(1987).
 DR EMBL; X16486; CAA34504.1; -.
 DR HSSP; P20693; 1HLJ.
 DR INTERPRO: IPR000436; -.
 DR INTERPRO: IPR001304; -.
 DR INTERPRO: IPR002353; -.
 DR PFAM; PF00059; lectin_c; 1.
 DR PFAM; PF00084; sushi; 1.
 DR PRINTS; PR00356; ANTIREFREZELI.
 DR PROSITE; PS00615; C-TYPE-LECTIN_1; 1.
 DR PROSITE; PS50041; C-TYPE-LECTIN_2; 1.
 KW Proteoglycan.
 FT NON_TER 1
 SO SEQUENCE 719 AA; 74501 MW; 33DDBA14CA1AA962 CRC64;

Query Match 96.9%; Score 31; DB 6; Length 719;
 Best Local Similarity 83.3%; Pred. No. 2.5e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEPGV 6
 DB 16 GEPGV 21

RESULT 8
 ID 015598 PRELIMINARY; PRT; 4650 AA.
 AC 015598;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE TITIN (FRAGMENT).
 GN TITIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-HEART;
 RX MEDLINE-94008990; PubMed-8404852;
 RA Gautel M., Leonard K., Labail S.;
 RT "Phosphorylation of KSP motifs in the C-terminal region of titin in
 differentiating myoblasts."
 RL EMBO J. 12:3627-3634(1993).
 DR EMBL; X69490; CAA49245.1; -.

DR HSSP: P00518; 1PK.
 DR INTERPRO: IPR000719; -
 DR INTERPRO: IPR001245; -
 DR INTERPRO: IPR001777; -
 DR INTERPRO: IPR002016; -
 DR INTERPRO: IPR003006; -
 DR PFAM: PF00041; fn3; 16.
 DR PFAM: PF00047; 1g; 17.
 DR PFAM: PF00069; pkinase; 1.
 DR PRINTS: PR00014; FNTYPERIT.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS00435; PEROXIDASE_1; UNKNOWN_1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR NON_TER 1
 SQ SEQUENCE 4650 AA; 522065 MW; C1P011FC313533DD CRC64;

Query Match 96.9%; Score 31; DB 4; Length 4650;
 Best Local Similarity 83.3%; Pred. NO. 1.8e+03;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEPAGV 6
 11111;
 Db 1757 GEPAGI 1762

RESULT 9
 ID 010466 PRELIMINARY; PRT; 26926 AA.
 AC 010466;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
 DE TITIN, HEART ISOFORM N2-B (EC 2.7.1.-) (CONNECTIN).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC TISSUE=HEART;
 RX MEDLINE=96026330; PubMed=7569978;
 RA Labelt S., Kolmer B.;
 RT "Titins: giant proteins in charge of muscle ultrastructure and elasticity.";
 RT Science 270:293-296(1995).
 RN [2]
 RN SEQUENCE OF 22277-25376 FROM N.A.
 RX MEDLINE=92258380; PubMed=1582406;
 RA Labelt S., Gautel M., Lakey A., Trinick J.;
 RT "Towards a molecular understanding of titin.";
 RT EMBO J. 11:1711-1716(1992).
 RN [3]
 RN SEQUENCE OF 1976-2014 FROM N.A.
 RA Labelt S.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RN CHARACTERIZATION.
 RX MEDLINE=95331314; PubMed=7607248;
 RA Gautel M., Castiglione-Morelli M.A., Pfuhl M., Motta A., Pastore A.;
 RT "A calmodulin-binding sequence in the C-terminus of human cardiac titin kinase.";
 RT Eur. J. Biochem. 230:752-759(1995).
 RL -1- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MUSCLE ASSEMBLY AND IN MAINTAINING THE STRUCTURAL INTEGRITY OF SARCOMERES. MAY HAVE PROTEIN KINASE ACTIVITY.
 CC -1- ALTERNATIVE PRODUCTS: A NUMBER OF FORMS OF THIS PROTEIN ARE PRODUCED BY ALTERNATIVE SPLICING WHICH DIFFER IN TISSUE DISTRIBUTION. DIFFERENT SIZE TRANSCRIPTS MAY ALSO EXIST WITHIN ANY ONE TISSUE. THE SEQUENCE SHOWN HERE IS THAT OF THE HEART ISOFORM N2-B.
 CC -1- TISSUE SPECIFICITY: MUSCLE-SPECIFIC.
 CC -1- SIMILARITY: TO THE CATALYTIC DOMAINS OF OTHER SERINE/THREONINE

CC KINASES.
 CC -1- SIMILARITY: BELONGS TO IMMUNOGLOBULIN SUPERFAMILY. CONTAINS 112
 CC IMMUNOGLOBULIN C2-LIKE DOMAINS AND 132 FIBRONECTIN TYPE III-LIKE
 CC DOMAINS.
 DR EMBL: X64698; CAA45939.1; -
 DR EMBL: X83270; CAA58243.1; -
 DR EMBL: X64697; CAA45938.1; -
 DR EMBL: X90568; CAA62188.1; -
 DR EMBL: X64699; CAA45940.1; -
 DR HSSP: P56276; ITLK.
 DR INTERPRO: IPR000129; -
 DR INTERPRO: IPR000282; -
 DR INTERPRO: IPR000577; -
 DR INTERPRO: IPR000719; -
 DR INTERPRO: IPR001245; -
 DR INTERPRO: IPR001777; -
 DR INTERPRO: IPR002016; -
 DR INTERPRO: IPR003006; -
 DR INTERPRO: IPR003015; -
 DR PFAM: PF00041; fn3; 132.
 DR PFAM: PF00047; 1g; 95.
 DR PFAM: PF00069; pkinase; 1.
 DR PRINTS: PR00014; FNTYPERIT.
 DR PRINTS: PR00726; LEXASERPTASE.
 DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS00290; IG_MHC; UNKNOWN_1.
 DR PROSITE: PS00435; PEROXIDASE_1; UNKNOWN_1.
 DR PROSITE: PS00933; EGGY_KINASES_1; UNKNOWN_1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Muscle protein; Cytoskeleton; Structural protein; Calmodulin-binding;
 KW Serine/threonine-protein kinase; Alternative splicing; Repeat;
 FT DOMAIN 1370 1389
 FT DOMAIN 4429 4614
 FT DOMAIN 24731 25070
 FT DOMAIN 25030 25056
 FT MOD_RES 1372 1372
 FT MOD_RES 1377 1377
 FT MOD_RES 1382 1382
 FT MOD_RES 1387 1387
 FT MOD_RES 26171 26171
 FT MOD_RES 26178 26178
 FT MOD_RES 26184 26184
 FT MOD_RES 26190 26190
 FT MOD_RES 22277 22277
 FT CONFLICT 22449 22449
 FT CONFLICT 22454 22454
 FT CONFLICT 23324 23324
 SQ SEQUENCE 26926 AA; 2993428 MW; D5EBCD3254DF5523 CRC64;

Query Match 96.9%; Score 31; DB 4; Length 26926;
 Best Local Similarity 83.3%; Pred. NO. 1.1e+04;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEPAGV 6
 11111;
 Db 24033 GEPAGI 24038

RESULT 10
 ID P95499 PRELIMINARY; PRT; 277 AA.
 AC P95499;
 DT 01-MAY-1997 (TReMBLrel. 03, Created)
 DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)
 DT 01-JAN-1999 (TReMBLrel. 09, Last annotation update)
 DE HDPH PROTEIN.
 GN HDPH.
 OS Pseudomonas hydrogenvora.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.

OX NCBI_TaxID=39439;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9-5;
 RA Okazaki M.;
 RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9-5;
 RX MEDLINE=96105206; PubMed=8529899;
 RA Ohtsuki T., Okazaki T., Endo Y., Kita Y., Shimosaka M., Okazaki M.;
 RT "Cloning and sequencing of the membrane-bound hydrogenase-encoding
 RT genes (hups and hupL) from Pseudomonas hydrogenovora.";
 RL Gene 166:89-93(1995).
 DR EMBL: D86986; BAA13228.1;
 SO SEQUENCE 277 AA; 30535 MW; 9ACFD92763C4BED9 CRC64;

Query Match 90.6%; Score 29; DB 2; Length 277;
 Best Local Similarity 83.3%; Pred. No. 2.4e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEPGV 6
 Db 130 GDAPGV 135

RESULT 11
 O9PC20 PRELIMINARY; PRT; 301 AA.
 AC O9PC20;
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
 DE HALOALKANE DEHALOGENASE.
 GN XHP965.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Araya J.E., Bata G.S., Baptista C.S.,
 RA Alvares R., Alves L.M.C., Araya J.E., Bata G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Doriry H.,
 RA Facinanci A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hohnsels J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marinho C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A., Paris A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Quagiro R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terezi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zait M., Zeldanis J., Zetubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
 RL Nature 406:151-157(2000).
 DR EMBL: AE004016; AAF84767.1; -

DR INTERPRO: IPR000073; -
 DR INTERPRO: IPR000379; -
 DR INTERPRO: IPR000639; -
 DR PFAM: PF00561; abhydrolase.1.
 DR PRINTS: PR00412; EPOXHYDROLASE.
 SO SEQUENCE 301 AA; 33882 MW; AC4CD590BD4D56B4 CRC64;

Query Match 90.6%; Score 29; DB 2; Length 301;
 Best Local Similarity 83.3%; Pred. No. 2.6e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEPGV 6
 Db 75 GDAPGV 80

RESULT 12
 O87803 PRELIMINARY; PRT; 332 AA.
 AC O87803;
 DT 01-NOV-1998 (TReMBLrel. 08, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
 DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
 DE OXIDOREDUCTASE.
 GN TOUF.
 OS Pseudomonas stutzeri (Pseudomonas perfectomarina).
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=316;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OX1;
 RX MEDLINE=98432776; PubMed=9758777;
 RA Bertoni G., Martino M., Galli E., Barbieri P.;
 RT "Analysis of the gene cluster encoding toluene/o-xylene monooxygenase
 RT from Pseudomonas stutzeri OX1.";
 RL Appl. Environ. Microbiol. 64:3626-3632(1998).
 DR EMBL: AJ005663; CAA06659.1; -
 DR HSSP: P11053; IPRD.
 DR INTERPRO: IPR000564; -
 DR INTERPRO: IPR001041; -
 DR INTERPRO: IPR001433; -
 DR PFAM: PF00111; fer2.1.
 DR PFAM: PF00175; oxidored.fad.1.
 DR PROSITE: PS00197; 2FE2S_FERRDOXIN.1.
 SO SEQUENCE 332 AA; 36560 MW; 668706D1182DD8F0 CRC64;

Query Match 90.6%; Score 29; DB 2; Length 332;
 Best Local Similarity 83.3%; Pred. No. 2.9e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEPGV 6
 Db 59 GEPGL 64

RESULT 13
 O9VLV6 PRELIMINARY; PRT; 369 AA.
 AC O9VLV6;
 DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
 DE CG1102 PROTEIN.
 GN CG1102.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Anagnostou C.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA Anagnostou P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sultun G.G., Wotman J.R., Tandel M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agabayan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintinas S.,
 RA Borokova D., Botchan M.R., Bouck J., Brooksstein P., Brotler P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeigwe C.,
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Kutsch K.A.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacלב J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relneer R., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skipski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhao M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003619; AAF52575.1;
 DR FLYBASE: FBgn0031961; CG7102.
 SQ SEQUENCE 369 AA; 40698 MW; 479A76A2E69711A1 CRC64;

Query Match 90.6%; Score 29; DB 5; Length 369;
 Best Local Similarity 83.3%; Pred. No. 3.2e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEPGV 6
 I:||||
 Db 355 GQAPGV 360

RESULT 14
 P95726 PRELIMINARY; PRT; 413 AA.
 AC P95726;
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
 DE SPAA GENE.
 GN SPAA.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97080529; PubMed=8921874;
 RX Schneider D., Bruton C.J., Chater K.F.;

RT "Characterization of spaa, a Streptomyces coelicolor gene homologous
 RT to a gene involved in sensing starvation in *Escherichia coli*.";
 RL Gene 117:243-251(1996).
 DR EMBL: X94190; CA63900.1;
 DR INTERPRO: IPR001354;
 DR PRAM: PF01188; MR_ME_1.
 DR PROSITE: PS00908; MR_ME_1;
 SQ SEQUENCE 413 AA; 45206 MW; 7EDC6183BC8DE36F CRC64;

Query Match 90.6%; Score 29; DB 2; Length 413;
 Best Local Similarity 83.3%; Pred. No. 3.6e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEPGV 6
 I:||||
 Db 215 GEPGL 220

RESULT 15
 O87843 PRELIMINARY; PRT; 443 AA.
 AC O87843;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE HYPOTHEICAL 48.1 KDA PROTEIN.
 GN SCBA6_09C.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Seeger K.J., Harris D.;
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Krieser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AL031013; CA119781.1;
 KW Hypothetical protein.
 SQ SEQUENCE 443 AA; 48140 MW; 005635215660881 CRC64;

Query Match 90.6%; Score 29; DB 2; Length 443;
 Best Local Similarity 83.3%; Pred. No. 3.9e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEPGV 6
 I:||||
 Db 215 GEPGL 220

Search completed: March 6, 2001, 12:52:59
 Job time: 299 sec

THIS PAGE BLANK (USPTO)

PI Dehoff BS, Kuhstoss SA, Rostock PR, Sutton KL;
 DX WPI; 1997-418046/39.
 DR N-PSDB; T80413.
 XX
 PT DNA encoding Streptomyces fradiae tylactone synthase domain - for
 PT production of tylosin-related polyketide compounds
 XX
 PS Claim 9; Pages 98-103; 220pp; English.
 CC W22601-W22605 represent proteins encoded by the tylactone synthase gene
 CC cluster of the invention. The gene cluster is also referred to as the
 CC tylG gene, and was isolated from Streptomyces fradiae. These sequences
 CC are multifunctional proteins which direct the synthesis of the polyketide
 CC tylactone, isolated from Streptomyces fradiae. Tylactone is the basic
 CC building block of the antibiotic tylosin. The DNA sequence can be
 CC modified so as to alter the type of carboxylic acids incorporated, the
 CC number of carboxylic acids incorporated and/or the post-condensation
 CC reactions performed, thereby resulting in novel tylosin-related
 CC polyketides.
 CC
 XX
 SQ Sequence 1611 AA;

Query Match 74.0%; Score 37; DB 18; Length 1611;
 Best Local Similarity 80.0%; Pred. No. 2.5e+02;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VFAGAAAG 10
 :|||||
 Db 1415 lfaagfaag 1424

RESULT 2

ID R22383
 R22383 standard; Protein; 194 AA.

AC R22383;

DT 17-AUG-1992 (first entry)

DE Antigen tc-239.

KW Oocysts; chicken antiserum; sporozoite.

OS Elmeria tenella.

PN W09204461-A.

PD 19-MAR-1992.

PF 05-SEP-1991; 91WO-US06431.

PR 12-SEP-1990; 90US-0501694.

PA (GENE-) GENEX CORP.

PA (FARH) HOECHST AG.

PI Jacobson JW, Strausberg RL, Wilson SD, Pope SH, Strausberg SL;
 PI Raether W;

DR WPI; 1992-114366/14.

DR N-PSDB; Q23083.

PT Vaccine against avian coccidiosis - comprising recombinant
 PT Elmeria antigen mc-4c, mc-5c or mc-30c gene, etc., or

PT microorganisms expressing them

PS Claim 14; Page 65 + Fig 14; 94pp; English.

CC To identify antigens of E. tenella, expression libraries were
 CC prepd. in the lambda vector, lambda gtl1, using cDNA prepd. from
 CC polyA mRNA isolated from E. tenella oocysts. The cDNA expression

CC library was screened with chicken antiserum raised against E.
 CC tenella sporozoites. The library to be screened was plated on
 CC a host that allows lysis and plaque formation. Following induction
 CC of the antigens encoded by the phage, the plaques were transferred
 CC to nitrocellulose filters. Positive phage were identified after
 CC screening with the chicken anti-E. tenella sporozoite antiserum.
 CC The cDNA inserts from the positive clones were cloned into
 CC bacteriophage M13 and subjected to sequence analysis.
 CC E. tenella antigen tc-239 (20 kd) was identified.
 CC
 XX
 SQ Sequence 194 AA;

Query Match 72.0%; Score 36; DB 13; Length 194;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 AAGAAAG 10
 :|||||
 Db 144 aagaaag 151

RESULT 3

ID W01142
 W01142 standard; Protein; 280 AA.

AC W01142;

DT 18-DEC-1996 (first entry)

DE TriPA protein.

KW TriPA; trpB; tryptophan synthase; L-tryptophan; essential; food;
 KW drug; stockbreeding; coryneform bacteria; transform.

OS Brevibacterium flavum.

PN JP08070871-A.

PD 19-MAR-1996.

PF 05-SEP-1994; 94JP-0211319.

PR 05-SEP-1994; 94JP-0211319.

PA (MITU) MITSUBISHI CHEM CORP.

DR WPI; 1996-203151/21.

DR N-PSDB; T40802.

PT DNA contg. a gene encoding tryptophan synthase - useful for
 PT L-tryptophan production in Coryneform bacteria

PS Claim 1; Page 4-8; 8pp; Japanese.

CC The present sequence that of the TriPA protein, part of a tryptophan
 CC synthase. Coryneform bacteria transformed with DNA (T40802) encoding it

CC and TriPA protein, express the tryptophan synthase which is useful for the

CC prodn. of L-tryptophan, an essential amino acid useful in foods, drugs

CC and stockbreeding.

XX
 SQ Sequence 280 AA;

Query Match 72.0%; Score 36; DB 17; Length 280;
 Best Local Similarity 70.0%; Pred. No. 66;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 FAAGAAAGV 11
 :|:|
 Db 139 fsaagaaag 148

```

RESULT 4
R05110 ID R05110 standard; protein; 379 AA.
XX AC
XX R05110;
XX DT 12-JUL-1990 (first entry)
XX DE Pella Pectin lyase recombinant gene product.
XX KW Pectin lyase; PLA; expression system; interferon.
XX KW ds.
XX OS Aspergillus niger.
XX PN EP353188-A.
XX PD 31-JAN-1990.
XX PF 19-JUL-1989; 89EP-0810545.
XX PR 28-JUL-1988; 88GB-0018046.
XX PR 26-JUN-1989; 89GB-0014666.
XX PA (CIBA ) CIBA GEIGY AG.
XX PI Heim J, Meyhack B, Visser J;
XX PT used to produce pectin lyase(s) PLA, PLB, PLC, PLE or PLF -
XX PT hybrid vectors expressing foreign genes.
XX PS Claim 3; Fig 10; 63pp; English.
XX CC Pella produced from hybrid vectors expressing the pectin lyase gene
XX CC and/or hybrid vectors expressing foreign genes eg. interferon
XX CC within a filamentous fungal expression system esp. Aspergillus niger.
XX SQ Sequence 379 AA;

Query Match 72.0%; Score 36; DB 11; Length 379;
Best Local Similarity 72.7%; Pred. No. 89;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 VFAGAAAGV 11
Db 13 vfagsaaavgv 23

RESULT 5
W32439 ID W32439 standard; Protein; 107 AA.
XX AC
XX W32439;
XX DT 08-JAN-1998 (first entry)
XX DE Mycobacterium tuberculosis antigen TbrAb.
XX KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
XX KW skin testing; M.tuberculosis.
XX OS Mycobacterium tuberculosis.
XX FH Key Location/Qualifiers
XX FT Misc-difference 60
XX FT Misc-difference 64 /note= "Any amino acid"
XX FT Misc-difference 64 /note= "Any amino acid"

```

```

FT Misc-difference 73 /note= "Any amino acid"
FT Misc-difference 77 /note= "Any amino acid"
FT Misc-difference 79 /note= "Any amino acid"
FT Misc-difference 86 /note= "Any amino acid"
FT Misc-difference 99 /note= "Any amino acid"
FT Misc-difference 102 /note= "Any amino acid"
FT Misc-difference 103 /note= "Any amino acid"
FT Misc-difference 106 /note= "Any amino acid"
FT Misc-difference 106 /note= "Any amino acid"
XX WO9709428-A2.
XX PD 13-MAR-1997.
XX PF 30-AUG-1996; 96WO-US14674.
XX PR 12-JUL-1996; 96US-0680574.
XX PR 01-SEP-1995; 95US-0523436.
XX PR 22-SEP-1995; 95US-0533634.
XX PR 22-MAR-1996; 96US-0620874.
XX PR 05-JUN-1996; 96US-0659683.
XX PA (CORI-) CORIXA CORP.
XX PI Campos-neto A, Dillon DC, Houghton R, Read SG, Skeiky YAW;
XX PI Twardzik DR, Vedvick TH;
XX DR WPI; 1997-192903/17.
XX DR N-PSDB; T91481.
XX PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
XX PT useful in vaccines for prevention or treatment of tuberculosis, also
XX PT for diagnosis
XX PS Example 3; Page 120; 168pp; English.
XX CC A new immunogenic polypeptide has been developed comprising an
XX CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
XX CC its variant differing only in conservative substitutions and/or
XX CC modifications). The present sequence represents a M.tuberculosis
XX CC antigen, TbrAb. The immunogenic protein, and fusion proteins
XX CC containing one or more of the proteins or one of the proteins plus
XX CC ESAT-6, are useful in vaccines, preferably when formulated with a
XX CC non-specific adjuvant, to induce an immune response against
XX CC M.tuberculosis (for treatment or prevention).
XX SQ Sequence 107 AA;

Query Match 70.0%; Score 35; DB 18; Length 107;
Best Local Similarity 88.9%; Pred. No. 37;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAGAAAGV 11
Db 25 aigaaaavgv 33

RESULT 6
W32371 ID W32371 standard; Protein; 107 AA.
XX AC
XX W32371;
XX DT 13-JAN-1998 (first entry)
XX

```

DE Mycobacterium tuberculosis antigen TbaB.
 XX
 KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
 XX skin testing; M.tuberculosis.
 XX
 OS Mycobacterium tuberculosis.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 60 /note= "Any amino acid"
 FT Misc-difference 64 /note= "Any amino acid"
 FT Misc-difference 73 /note= "Any amino acid"
 FT Misc-difference 77 /note= "Any amino acid"
 FT Misc-difference 79 /note= "Any amino acid"
 FT Misc-difference 86 /note= "Any amino acid"
 FT Misc-difference 99 /note= "Any amino acid"
 FT Misc-difference 102 /note= "Any amino acid"
 FT Misc-difference 103 /note= "Any amino acid"
 FT Misc-difference 106 /note= "Any amino acid"
 FT
 XX WO9709429-A2.
 XX
 PN 13-MAR-1997.
 XX
 XX 30-AUG-1996; 96WO-US14675.
 XX
 PR 12-JUL-1996; 96US-0680573.
 PR 01-SEP-1995; 95US-0523435.
 PR 22-SEP-1995; 95US-0532136.
 PR 22-MAR-1996; 96US-0620280.
 PR 05-JUN-1996; 96US-0658800.
 XX
 XX (CORI-) CORIXA CORP.
 XX
 PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;
 PI Twardzik DR, Vedvick TH;
 XX
 XX WPI: 1997-192904/17.
 DR N-PSDB; T91418.
 XX
 XX New immunogenic polypeptide(s) from soluble M. tuberculosis antigens
 PT - useful for diagnosis of M. tuberculosis infection
 XX
 XX Example 3; Page 131-132; 190pp; English.
 XX
 CC A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
 CC its variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis
 CC antigen, TbaB. The immunogenic polypeptide can be used to diagnose
 CC M.tuberculosis infection by forming complexes with specific
 CC antibodies in the sample. Fragments of DNA encoding the immunogenic
 CC polypeptide can be used as diagnostic primers or probes and agents
 CC that bind to the antigen, especially monoclonal antibodies or
 CC equivalent polyclonal antibodies, are also used for diagnosis.
 XX
 XX Sequence 107 AA;
 SQ
 Query Match 70.0%; Score 35; DB 18; Length 107;
 Best Local Similarity 88.9%; Pred. No. 37;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 3 AAGAAAGV 11

Db 25 aigaaaagv 33
 | |||||
 RESULT 7
 W81674
 ID W81674 standard; Protein; 107 AA.
 XX
 XX W81674;
 XX
 DT 27-JAN-1999 (first entry)
 XX
 DE M. tuberculosis immunogenic polypeptide TbaB.
 XX
 KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
 KW vaccine; pharmaceutical; infection; diagnosis.
 XX
 OS Mycobacterium tuberculosis.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 60 /label= unknown
 FT Misc-difference 64 /label= unknown
 FT Misc-difference 73 /label= unknown
 FT Misc-difference 77 /label= unknown
 FT Misc-difference 79 /label= unknown
 FT Misc-difference 86 /label= unknown
 FT Misc-difference 99 /label= unknown
 FT Misc-difference 102 /label= unknown
 FT Misc-difference 103 /label= unknown
 FT Misc-difference 106 /label= unknown
 FT
 XX WO9816646-A2.
 XX
 PN 23-APR-1998.
 XX
 XX 07-OCT-1997; 97WO-US18293.
 XX
 PR 13-MAR-1997; 97US-0818112.
 PR 11-OCT-1996; 96US-0730510.
 XX
 XX (CORI-) CORIXA CORP.
 XX
 XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX
 XX WPI: 1998-261042/23.
 DR N-PSDB; V64467.
 XX
 XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
 PT to develop products for the detection of M. tuberculosis infection
 PT and for diagnosis, treatment and prevention of tuberculosis
 XX
 PS Example 3; Page 115; 230pp; English.
 XX
 CC This sequence represents an immunogenic portion of a soluble
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method
 CC for inducing protective immunity against tuberculosis (TB). This sequence
 CC can be formulated into vaccines and/or pharmaceutical compositions for
 CC immunising against M. tuberculosis infection or may be used for the
 CC diagnosis of tuberculosis.
 XX
 SQ Sequence 107 AA;

Query Match 70.0%; Score 35; DB 19; Length 107;
 Best Local Similarity 88.9%; Pred. NO. 37;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 AAGAAAAGV 11
 | | | | | | | |
 Db 25 aigaaaagv 33

RESULT 8

W64311
 ID W64311 standard; Protein; 107 AA.

AC W64311;

DT 09-NOV-1998 (first entry)

DE Mycobacterium tuberculosis antigen TbrAB.

KW Tuberculosis; infection; diagnosis; antigen; TbrAB.

OS Mycobacterium tuberculosis strain H37Ra.

FH Key Location/Qualifiers

FT Misc-difference 40 /note= "encoded by GGN"

FT Misc-difference 60 /note= "encoded by GAN"

FT Misc-difference 64 /note= "encoded by GAN"

FT Misc-difference 73 /note= "encoded by GAN"

FT Misc-difference 77 /note= "encoded by AGN"

FT Misc-difference 79 /note= "encoded by AGN"

FT Misc-difference 86 /note= "encoded by GNC"

FT Misc-difference 89 /note= "encoded by GNG"

FT Misc-difference 94 /note= "encoded by GGN"

FT Misc-difference 95 /note= "encoded by GGN"

FT Misc-difference 97 /note= "encoded by GGN"

FT Misc-difference 99 /note= "encoded by GGN"

FT Misc-difference 101 /note= "encoded by ANC"

FT Misc-difference 102 /note= "encoded by GGN"

FT Misc-difference 103 /note= "encoded by GNG"

FT Misc-difference 104 /note= "encoded by NAT"

FT Misc-difference 106 /note= "encoded by GGN"

FT Misc-difference 106 /note= "encoded by NCA"

XX WO9816645-A2.

XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US18214.

XX 13-MAR-1997; 97US-0818111.

XX 11-OCT-1996; 96US-0729622.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX

DR WPI; 1998-251292/22.

DR N-PSDB; V44359.

XX

PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used
 PT to develop products for the detection of M. tuberculosis infection
 PT and diagnosis of tuberculosis

XX

PS Example 3; Page 120; 250pp; English.

XX

CC This polypeptide comprises Mycobacterium tuberculosis soluble
 CC antigen TbrAB. It is encoded by a DNA sequence (see V44359)
 CC isolated from a M. tuberculosis strain H37Ra expression library
 CC with rabbit anti-sera raised against M. tuberculosis supernatant.
 CC The invention relates to compositions and methods for diagnosing
 CC tuberculosis. It provides polypeptides (see W64291-W64379)
 CC comprising an antigenic portion of a soluble M. tuberculosis
 CC antigen, or an immunogenic portion of an M. tuberculosis antigen,
 CC as well as DNA sequences encoding such polypeptides, recombinant
 CC expression vectors and transformed or transfected host cells. Also
 CC claimed are methods and diagnostic kits for detecting M.
 CC tuberculosis infection in a patient using the above polypeptides,
 CC antibodies, or oligonucleotide probes and primers, for the
 CC diagnosis of tuberculosis.

XX

SQ Sequence 107 AA;

Query Match 70.0%; Score 35; DB 19; Length 107;

Best Local Similarity 88.9%; Pred. NO. 37;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 AAGAAAAGV 11

| | | | | | | |

Db 25 aigaaaagv 33

;

RESULT 9

Y39113

ID Y39113 standard; Protein; 107 AA.

XX

AC Y39113;

XX

DT 05-NOV-1999 (first entry)

XX

DE M. tuberculosis antigen TbrAB amino acid sequence.

XX

KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;

KW immunotherapy; diagnosis; immunisation; vaccine; infection;

KW immune response; skin test.

XX

OS Mycobacterium tuberculosis.

XX

PN WO9942076-A2.

XX

PD 26-AUG-1999.

XX

PF 17-FEB-1999; 99WO-US03268.

XX

PR 05-MAY-1998; 98US-0072967.

XX

PR 18-FEB-1998; 98US-0025197.

XX

PA (CORI-) CORIXA CORP.

XX

XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;

XX Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1999-527409/44.

XX N-PSDB; Z19269.

XX New antigens from Mycobacterium tuberculosis useful in diagnostic
 XX skin tests and protective or therapeutic vaccines or compositions

PS Example 3; Page 111; 299pp; English.

CC The present invention describes polypeptides comprising an immunogenic
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
 CC are vaccines and fusion protein containing M. tuberculosis Ag's.
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
 CC other polypeptides fragments, can be used in pharmaceutical compositions
 CC or vaccines to generate a protective or therapeutic immune response to
 CC M. tuberculosis and as reagents in skin tests for diagnosis of
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
 CC by, T, B or natural killer cells and/or macrophages in
 CC tuberculosis-immune subjects. Z19249 to Z19450 and Y39083 to Y39225
 CC are used in the exemplification of the present invention.

XX Sequence 107 AA;

Query Match 70.0%; Score 35; DB 20; Length 107;
 Best Local Similarity 88.9%; Pred. No. 37;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAGAAAAGV 11
 | | | | |
 Db 25 aigaaaagv 33

RESULT 10

Y38976
 ID Y38976 standard; Protein; 107 AA.

XX

AC Y38976;

XX 05-NOV-1999 (first entry)

XX M. tuberculosis recombinant antigen protein TbrAb.

XX Antigen; diagnosis; detection; infection; antibody; immunisation;
 KW vaccine; immunity.

XX Mycobacterium tuberculosis.

XX WO9942118-A2.

XX 26-AUG-1999.

XX 17-FEB-1999; 99WO-US03265.

XX 05-MAY-1998; 98US-0072596.

XX 18-FEB-1998; 98US-0024753.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;

PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1999-527416/44.

XX N-PSDB; Z19057.

XX New polypeptide comprising antigenic portions of M. tuberculosis

XX Example 3; Page 156; 323pp; English.

XX This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against
 CC M. tuberculosis infection. The new detection methods are needed as
 CC current vaccination strategies do not provide 100% immunity.

XX Sequence 107 AA;

SQ

Query Match 70.0%; Score 35; DB 20; Length 107;
 Best Local Similarity 88.9%; Pred. No. 37;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAGAAAAGV 11
 | | | | |
 Db 25 aigaaaagv 33

RESULT 11

W32465

ID W32465 standard; Protein; 168 AA.

XX W32465;

XX 09-JAN-1998 (first entry)

XX Mycobacterium tuberculosis antigen DPAS.

XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
 KW skin testing; M.tuberculosis.

XX Mycobacterium tuberculosis.

XX WO9709428-A2.

XX 13-MAR-1997.

XX 30-AUG-1996; 96WO-US14674.

XX 12-JUL-1996; 96US-0680574.

XX 01-SEP-1995; 95US-0523436.

XX 22-SEP-1995; 95US-0536334.

XX 22-MAR-1996; 96US-0620874.

XX 05-JUN-1996; 96US-0659683.

XX (CORI-) CORIXA CORP.

XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;

PI Twardzik DR, Vedvick TH;

XX WPI; 1997-192903/17.

XX N-PSDB; T91487.

XX New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
 PT useful in vaccines for prevention or treatment of tuberculosis, also
 PT for diagnosis

XX Disclosure; Page 132-133; 168pp; English.

XX A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
 CC its variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis
 CC antigen, DPAS. The immunogenic protein, and fusion proteins
 CC containing one or more of the proteins or one of the proteins plus
 CC ESAT-6, are useful in vaccines, preferably when formulated with a
 CC non-specific adjuvant, to induce an immune response against
 CC M.tuberculosis (for treatment or prevention).

XX Sequence 168 AA;

Query Match 70.0%; Score 35; DB 18; Length 168;
 Best Local Similarity 88.9%; Pred. No. 58;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAGAAAAGV 11
 | | | | |
 Db 16 aigaaaagv 24

RESULT 12

W32393
ID W32393 standard; Protein; 168 AA.
XX
AC W32393;
XX
DT 12-JAN-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen DPAS.
XX
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW skin testing; M.tuberculosis.
XX
OS Mycobacterium tuberculosis.
XX
PN W09709429-A2.
XX
PD 13-MAR-1997.
XX
PF 30-AUG-1996; 96WO-US14675.
XX
PR 12-JUL-1996; 96US-0680573.
PR 01-SEP-1995; 95US-0523435.
PR 22-SEP-1995; 95US-0532136.
PR 22-MAR-1996; 96US-0620280.
PR 05-JUN-1996; 96US-0658800.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;
PI Twardzik DR, Vedvick TH;
XX
DR WPI; 1997-192904/17.
XX
XX New immunogenic polypeptide(s) from soluble M. tuberculosis antigens
PT - useful for diagnosis of M. tuberculosis infection
XX
PS Claim 3; Page 142; 190pp; English.
XX
CC A new immunogenic polypeptide has been developed comprising an
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
CC its variant differing only in conservative substitutions and/or
CC modifications). The present sequence represents a specifically claimed
CC M.tuberculosis antigen, DPAS (in the specification claim 3 specifies
CC the DNA sequences however the present sequence represents the protein
CC sequence, which can be compared with W32465 from a similar
CC specification). The immunogenic polypeptide can be used to diagnose
CC M.tuberculosis infection by forming complexes with specific antibodies
CC in the sample. Fragments of DNA encoding the immunogenic polypeptide
CC can be used as diagnostic primers, or probes and agents that bind to
CC the antigen, especially monoclonal antibodies or equivalent polyclonal
CC antibodies, are also used for diagnosis.
CC N.B. The present sequence represents the protein sequence, not the
CC DNA as specified.
XX
SQ Sequence 168 AA;

Query Match 70.0%; Score 35; DB 18; Length 168;
Best Local Similarity 88.9%; Pred. No. 58;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 AAGAAAGV 11
I | | | | |
Db 16 aigaaaagv 24

RESULT 13
W81699
ID W81699 standard; Protein; 168 AA.
XX
AC W81699;
XX
DT 27-JAN-1999 (first entry)

XX M. tuberculosis immunogenic polypeptide DPAS.
DE
XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW vaccine; pharmaceutical; infection; diagnosis.
XX
OS Mycobacterium tuberculosis.
XX
PN W09816646-A2.
XX
PD 23-APR-1998.
XX
PF 07-OCT-1997; 97WO-US18293.
XX
PR 13-MAR-1997; 97US-0818112.
PR 11-OCT-1996; 96US-0730510.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI; 1998-261042/23.
DR N-PSDB; V64499.
XX
PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and for diagnosis, treatment and prevention of tuberculosis
PS Disclosure; Page 124; 230pp; English.
XX
CC This sequence represents an immunogenic portion of a soluble
CC Mycobacterium tuberculosis (MT) antigen which can be used in a method
CC for inducing protective immunity against tuberculosis (TB). This
CC sequence can be formulated into vaccines and/or pharmaceutical
CC compositions for immunising against M. tuberculosis infection or may
CC be used for the diagnosis of tuberculosis.
XX
SQ Sequence 168 AA;

Query Match 70.0%; Score 35; DB 19; Length 168;
Best Local Similarity 88.9%; Pred. No. 58;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 AAGAAAGV 11
I | | | | |
Db 16 aigaaaagv 24

RESULT 14
W64332
ID W64332 standard; Protein; 168 AA.
XX
AC W64332;
XX
DT 09-NOV-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen DPAS.
XX
KW Tuberculosis; infection; diagnosis; antigen; DPAS.
XX
OS Mycobacterium tuberculosis strain H37Rv.
XX
PN W09816645-A2.
XX
PD 23-APR-1998.
XX
PF 07-OCT-1997; 97WO-US18214.
XX
PR 13-MAR-1997; 97US-0818111.
PR 11-OCT-1996; 96US-0729622.
XX

PA (CORI-) CORIXA CORP.
 XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX WPI: 1998-251292/22.
 DR N-PSDB: V44391.
 XX
 PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used
 PT to develop products for the detection of M. tuberculosis infection
 PT and diagnosis of tuberculosis
 XX
 PS Example 1; Page 128; 250pp; English.
 XX
 CC This polypeptide comprises the Mycobacterium tuberculosis antigen
 CC DPAS. A DNA sequence (see V44391) coding for DPAS was isolated
 CC from a M. tuberculosis genomic library. An N-terminal peptide
 CC (see W64346) of DPAS was shown to induce proliferation and
 CC interferon-gamma production in peripheral blood mononuclear cells.
 CC DPAS shows sequence homology to a Mycobacterium leprae protein.
 CC The invention relates to compositions and methods for diagnosing
 CC tuberculosis. It provides polypeptides (see W64291-W64379)
 CC comprising an antigenic portion of a soluble M. tuberculosis
 CC antigen, or an immunogenic portion of an M. tuberculosis antigen,
 CC as well as DNA sequences encoding such polypeptides, recombinant
 CC expression vectors and transformed or transfected host cells.
 CC Also claimed are methods and diagnostic kits for detecting M.
 CC tuberculosis infection in a patient using these polypeptides,
 CC antibodies or oligonucleotide probes and primers, for the diagnosis
 CC of tuberculosis.
 XX
 SQ Sequence 168 AA;

Query Match 70.0%; Score 35; DB 19; Length 168;
 Best Local Similarity 88.9%; Pred. No. 58;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAGAAAAGV 11
 | | | | | | |
 Db 16 aigaaaagv 24

RESULT 15

Y39129
 ID Y39129 standard; Protein; 168 AA.

XX Y39129;
 AC Y39129;

XX 05-NOV-1999 (first entry)

XX M. tuberculosis antigen DPAS amino acid sequence.

XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;
 KW immune response; skin test.

XX Mycobacterium tuberculosis.

OS Mycobacterium tuberculosis.

PN WO9942076-A2.

XX 26-AUG-1999.

XX 17-FEB-1999; 99WO-US03268.

XX 05-MAY-1998; 98US-0072967.

XX 18-FEB-1998; 98US-0025197.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX

DR WPI: 1999-527409/44.
 DR N-PSDB: Z19301.
 XX
 PT New antigens from Mycobacterium tuberculosis useful in diagnostic
 PT skin tests and protective or therapeutic vaccines or compositions
 XX
 PS Disclosure; Page 118-119; 299pp; English.
 XX
 CC The present invention describes polypeptides comprising an immunogenic
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
 CC are vaccines and fusion protein containing M. tuberculosis Ag s.
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
 CC other polypeptides fragments, can be used in pharmaceutical compositions
 CC or vaccines to generate a protective or therapeutic immune response to
 CC M. tuberculosis and as reagents in skin tests for diagnosis of
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
 CC by, T, B or natural killer cells and/or macrophages in
 CC tuberculosis-immune subjects. Z19249 to Z19460 and Y39083 to Y39225
 CC are used in the exemplification of the present invention.
 XX
 SQ Sequence 168 AA;

Query Match 70.0%; Score 35; DB 20; Length 168;
 Best Local Similarity 88.9%; Pred. No. 58;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAGAAAAGV 11
 | | | | | | |
 Db 16 aigaaaagv 24

Search completed: March 6, 2001, 12:49:38
 Job time: 99 sec

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:53:51 ; Search time 57.76 Seconds
(without alignments)
3.420 Million cell updates/sec

Title: US-09-196-161d-8
Perfect score: 50
Sequence: 1 VFAAGAAAGV 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	74.0	1611	2	US-08-804-227C-5
2	36	72.0	379	1	US-07-723-002C-4
3	35	70.0	4551	3	US-09-320-878-1
4	34	68.0	54	1	US-08-209-747-6
5	34	68.0	54	1	US-08-458-298-6
6	34	68.0	155	1	US-08-209-747-15
7	34	68.0	155	1	US-08-458-298-15
8	34	68.0	407	2	US-08-765-875-2
9	34	68.0	407	2	US-08-765-875-6
10	34	68.0	407	3	US-08-795-671-2
11	34	68.0	407	3	US-08-795-671-6
12	34	68.0	832	1	US-08-209-747-2
13	34	68.0	832	1	US-08-458-298-2
14	33	66.0	49	3	US-08-398-633-18
15	33	66.0	49	3	US-08-480-070C-24
16	33	66.0	49	3	US-08-829-525-38
17	33	66.0	260	4	PCT-US96-03916-9
18	33	66.0	260	4	PCT-US96-03916-70
19	33	66.0	345	1	US-08-171-382-6
20	33	66.0	345	1	US-08-309-420-6
21	33	66.0	345	1	US-08-309-419-6
22	33	66.0	345	4	PCT-US95-11856-6
23	33	66.0	345	4	PCT-US95-11856-6
24	33	66.0	454	1	US-08-171-382-4
25	33	66.0	454	1	US-08-309-420-4
26	33	66.0	454	1	US-08-309-419-4
27	33	66.0	454	4	PCT-US95-11856-4
28	33	66.0	454	4	PCT-US95-11878-4

29 33 66.0 594 1 US-08-171-382-2 Sequence 2, Appli
30 33 66.0 594 1 US-08-309-420-2 Sequence 2, Appli
31 33 66.0 594 1 US-08-309-419-2 Sequence 2, Appli
32 33 66.0 594 4 PCT-US95-11856-2 Sequence 2, Appli
33 33 66.0 594 4 PCT-US95-11878-2 Sequence 2, Appli
34 33 66.0 951 1 US-08-162-809-2 Sequence 2, Appli
35 33 66.0 984 2 US-08-673-789-6 Sequence 6, Appli
36 33 66.0 3031 1 US-07-689-008-2 Sequence 2, Appli
37 32 64.0 8 2 US-08-612-785B-17 Sequence 17, Appli
38 32 64.0 14 1 US-08-556-823-3 Sequence 3, Appli
39 32 64.0 14 1 US-08-556-823-4 Sequence 4, Appli
40 32 64.0 18 3 US-09-053-617A-1 Sequence 1, Appli
41 32 64.0 19 1 US-08-556-823-5 Sequence 5, Appli
42 32 64.0 19 1 US-08-244-701B-3 Sequence 3, Appli
43 32 64.0 19 1 US-08-425-069-5 Sequence 5, Appli
44 32 64.0 28 1 US-08-425-069-5 Sequence 50, Appli
45 32 64.0 28 2 US-08-317-844B-50 Sequence 50, Appli

ALIGNMENTS

RESULT 1
US-08-804-227C-5
; Sequence 5, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1611 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-804-227C-5

Query Match 74.0%; Score 37; DB 2; Length 1611;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 VFAAGAAAG 10
Db 1415 LFAAGFAAG 1424

```

RESULT 2
US-07-723-002C-4
: Sequence 4, Application US/07723002C
: Patent No. 5447862
: GENERAL INFORMATION:
: APPLICANT: Helm, Jutta
: APPLICANT: Meyhack, Bernd
: APPLICANT: Gysler, Christof
: APPLICANT: Visser, Jacob
: APPLICANT: Kester, Hermanus Cornelis Maria
: TITLE OF INVENTION: No. 5447862el Expression System
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: CIBA-GEIGY Corporation
: STREET: 7 skyline Drive
: CITY: Hawthorne
: STATE: New York
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07723,002C
: FILING DATE: 28-JUN-1991
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 8702475
: FILING DATE: 04-FEB-1987
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 150,880
: FILING DATE: 29-JAN-1988
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 8818046.8
: FILING DATE: 28-JUL-1988
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 8914666.6
: FILING DATE: 26-JUN-1989
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 384,898
: FILING DATE: 24-JUL-1989
: ATTORNEY/AGENT INFORMATION:
: NAME: Foley, Shawn P.
: REGISTRATION NUMBER: 33,071
: REFERENCE/DOCKET NUMBER: 4-16317/+CIP
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (919)541-8615
: TELEFAX: (919)541-8689
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 379 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-07-723-002C-4

```

```

Query Match 72.0%; Score 36; DB 1; Length 379;
Best Local Similarity 72.7%; Pred. No. 59;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 VFAGAAAAGV 11
Db 13 VFAGAAAAGV 23

```

```

RESULT 3
US-09-320-878-1
: Sequence 1, Application US/09320878A
: Patent No. 6117659
: GENERAL INFORMATION:

```

```

: APPLICANT: ASHLEY, Gary
: APPLICANT: BETLACH, Melanie C.
: APPLICANT: BETLACH, Mary C.
: APPLICANT: MCDANIEL, Robert
: APPLICANT: TANG, Li
: TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
: FILE REFERENCE: 300622002120
: CURRENT APPLICATION NUMBER: US/09/320,878A
: CURRENT FILING DATE: 1999-05-27
: EARLIER APPLICATION NUMBER: CIP OF 09/141,908
: EARLIER FILING DATE: 1998-08-28
: EARLIER APPLICATION NUMBER: CIP OF 09/073,538
: EARLIER FILING DATE: 1998-05-06
: EARLIER APPLICATION NUMBER: CIP OF 08/846,247
: EARLIER FILING DATE: 1997-04-30
: EARLIER APPLICATION NUMBER: 60/119,139
: EARLIER FILING DATE: 1999-02-08
: EARLIER APPLICATION NUMBER: 60/100,880
: EARLIER FILING DATE: 1998-09-22
: EARLIER APPLICATION NUMBER: 60/087,080
: EARLIER FILING DATE: 1998-05-28
: NUMBER OF SEQ ID NOS: 34
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 1
: LENGTH: 4551
: TYPE: PRT
: ORGANISM: Streptomyces venezuelae
US-09-320-878-1

Query Match 70.0%; Score 35; DB 3; Length 4551;
Best Local Similarity 81.8%; Pred. No. 1e+03;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VFAGAAAAGV 11
Db 3767 VHAAGAAAARV 3777

RESULT 4
US-08-209-747-6
: Sequence 6, Application US/08209747
: Patent No. 5733771
: GENERAL INFORMATION:
: APPLICANT: Lewis, Randolph V.
: APPLICANT: Colgin, Mark
: TITLE OF INVENTION: cDNAs Encoding Minor Ampullate Spider
: NUMBER OF SEQUENCES: 56
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Birch, Stewart, Kolasch & Birch
: STREET: P.O. Box 747
: CITY: Falls Church
: STATE: Virginia
: COUNTRY: USA
: ZIP: 22040-3487
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/209,747
: FILING DATE: 14-MAR-1994
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: Murphy Jr., Gerald M.
: REGISTRATION NUMBER: 28,977
: REFERENCE/DOCKET NUMBER: 1447-104P
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-205-8000
: TELEFAX: 703-205-8050
: INFORMATION FOR SEQ ID NO: 6:

```


; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-209-747-6

Query Match 68.0%; Score 34; DB 1; Length 54;
Best Local Similarity 77.8%; pred. No. 18;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 FAAGAAAG 10
|||||
Db 27 FAAGAGGAG 35

RESULT 5

US-08-458-298-6
; Sequence 6, Application US/08458298
; Patent No. 5756677
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Colgin, Mark
; TITLE OF INVENTION: cDNAs Encoding Minor Ampullate Spider
; TITLE OF INVENTION: Silk Proteins
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-3487
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,298
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/209,747
; FILING DATE: 14-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-104P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-458-298-6

Query Match 68.0%; Score 34; DB 1; Length 54;
Best Local Similarity 77.8%; pred. No. 18;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 FAAGAAAG 10
|||||
Db 27 FAAGAGGAG 35

RESULT 6

US-08-209-747-15

; Sequence 15, Application US/08209747
; Patent No. 5733771
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Colgin, Mark
; TITLE OF INVENTION: cDNAs Encoding Minor Ampullate Spider
; TITLE OF INVENTION: Silk Proteins
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-3487
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/209,747
; FILING DATE: 14-MAR-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-104P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 155 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..155
; OTHER INFORMATION: /label= MISP.N.aa
; OTHER INFORMATION: /note= "amino-terminal sequence of mispl, see Fig.
; OTHER INFORMATION: 4"
US-08-209-747-15

Query Match 68.0%; Score 34; DB 1; Length 155;
Best Local Similarity 80.0%; pred. No. 51;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 VFAAGAAAG 10
| | | | |
Db 88 VAAAGAGAG 97

RESULT 7
US-08-458-298-15
; Sequence 15, Application US/08458298
; Patent No. 5756677
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Colgin, Mark
; TITLE OF INVENTION: cDNAs Encoding Minor Ampullate Spider
; TITLE OF INVENTION: Silk Proteins
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA

```

; ZIP: 22040-3487
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 02-JUN-1995
; APPLICATION NUMBER: US/08/458,298
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/209,747
; FILING DATE: 14-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-104P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 155 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..155
; OTHER INFORMATION: /label= MISP.N_aa
; OTHER INFORMATION: /note= "amino-terminal sequence of mispl, see Fig.
; US-08-458-298-15

Query Match 68.0%; Score 34; DB 1; Length 155;
Best Local Similarity 80.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VFAAGAAAG 10
Db 88 VAAAGAGAG 97

RESULT 8
US-08-765-875-2
; Sequence 2, Application US/08765875
; Patent No. 5914234
; GENERAL INFORMATION:
; APPLICANT: LEE, SE-JIN
; APPLICANT: MCPHERSON, ALEXANDRA C.
; TITLE OF INVENTION: GROWTH DIFFERENTIATION FACTOR-11
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SPENSLEY HORN JUBAS & LUBITZ
; STREET: 1880 CENTURY PARK EAST, FIFTH FLOOR
; CITY: LOS ANGELES
; STATE: CALIFORNIA
; COUNTRY: US
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/765,875
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/706,958
; FILING DATE:
; APPLICATION NUMBER: US/08/272,763
; FILING DATE: 08-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: TUMARKIN PH.D., LISA A.
; REGISTRATION NUMBER: P-38,347
; REFERENCE/DOCKET NUMBER: PD3641
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/455-5100
; TELEFAX: 619/455-5110
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 407 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:

```

```

; FILING DATE:
; APPLICATION NUMBER: US/08/272,763
; FILING DATE: 08-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: TUMARKIN PH.D., LISA A.
; REGISTRATION NUMBER: P-38,347
; REFERENCE/DOCKET NUMBER: PD3641
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/455-5100
; TELEFAX: 619/455-5110
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 407 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-765-875-2

Query Match 68.0%; Score 34; DB 2; Length 407;
Best Local Similarity 88.9%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAGAAAAGV 11
Db 35 AAAAAAGV 43

RESULT 9
US-08-765-875-6
; Sequence 6, Application US/08765875
; Patent No. 5914234
; GENERAL INFORMATION:
; APPLICANT: LEE, SE-JIN
; APPLICANT: MCPHERSON, ALEXANDRA C.
; TITLE OF INVENTION: GROWTH DIFFERENTIATION FACTOR-11
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SPENSLEY HORN JUBAS & LUBITZ
; STREET: 1880 CENTURY PARK EAST, FIFTH FLOOR
; CITY: LOS ANGELES
; STATE: CALIFORNIA
; COUNTRY: US
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/765,875
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/706,958
; FILING DATE:
; APPLICATION NUMBER: US/08/272,763
; FILING DATE: 08-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: TUMARKIN PH.D., LISA A.
; REGISTRATION NUMBER: P-38,347
; REFERENCE/DOCKET NUMBER: PD3641
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/455-5100
; TELEFAX: 619/455-5110
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 407 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:

```

```

; CLONE: GDF-11
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..407
US-08-765-875-6

```

```

Query Match      68.0%; Score 34; DB 2; Length 407;
Best Local Similarity 88.9%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 3 AAGAAAAGV 11
   || |||||
DB 35 AAAAAAGV 43

```

```

RESULT 10
US-08-795-671-2
; Sequence 2, Application US/08795671
; Patent No. 6008434
; GENERAL INFORMATION:
; APPLICANT: Se-Jin Lee and Alexandra McPherron
; TITLE OF INVENTION: GROWTH DIFFERENTIATION FACTOR-11
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,671
; FILING DATE: February 6, 1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: HAILE, PH.D., LISA A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07265/106001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 407 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-795-671-2

```

```

Query Match      68.0%; Score 34; DB 3; Length 407;
Best Local Similarity 88.9%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 3 AAGAAAAGV 11
   || |||||
DB 35 AAAAAAGV 43

```

```

RESULT 11
US-08-795-671-6
; Sequence 6, Application US/08795671
; Patent No. 6008434
; GENERAL INFORMATION:
; APPLICANT: Se-Jin Lee and Alexandra McPherron
; TITLE OF INVENTION: GROWTH DIFFERENTIATION FACTOR-11
; NUMBER OF SEQUENCES: 9

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,671
; FILING DATE: February 6, 1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: HAILE, PH.D., LISA A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07265/106001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 407 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: GDF-11
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..407
US-08-795-671-6

```

```

Query Match      68.0%; Score 34; DB 3; Length 407;
Best Local Similarity 88.9%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 3 AAGAAAAGV 11
   || |||||
DB 35 AAAAAAGV 43

```

```

RESULT 12
US-08-209-747-2
; Sequence 2, Application US/08209747
; Patent No. 573771
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Colgin, Mark
; TITLE OF INVENTION: cDNAs Encoding Minor Ampullate Spider
; TITLE OF INVENTION: Silk Proteins
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-3487
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/209,747
; FILING DATE: 14-MAR-1994
; CLASSIFICATION: 530

```

ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-104P
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 832 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: N. clavipes
TISSUE TYPE: minor ampullate gland
FEATURE:
NAME/KEY: CDS
LOCATION: 1..309
US-08-209-747-2

Query Match 68.0%; Score 34; DB 1; Length 832;
Best Local Similarity 80.0%; Pred. No. 2.7e+02;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 VFAAGAAAG 10
| | | | | | | |
Db 88 VAAAGAAG 97

RESULT 13

US-08-458-298-2
Sequence 2, Application US/08458298
Patent No. 5756677
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
TITLE OF INVENTION: CDNAs Encoding Minor Ampullate Spider
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-3487
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,298
FILING DATE: 02-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/209,747
FILING DATE: 14-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-104P
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 832 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: N. clavipes
TISSUE TYPE: minor ampullate gland
FEATURE:
NAME/KEY: CDS
LOCATION: 1..309
US-08-458-298-2

Query Match 68.0%; Score 34; DB 1; Length 832;
Best Local Similarity 80.0%; Pred. No. 2.7e+02;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 VFAAGAAAG 10
| | | | | | | |
Db 88 VAAAGAAG 97

RESULT 14

US-08-398-633-18
Sequence 18, Application US/08398633
Patent No. 6066322
GENERAL INFORMATION:
APPLICANT: Levinson, Douglas A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF IMMUNE DISORDERS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/398,633
FILING DATE: 03-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-011
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 49 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-398-633-18

Query Match 66.0%; Score 33; DB 3; Length 49;
Best Local Similarity 70.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VFAAGAAAG 10
| | | | | | | |

Db 3 VFLAGGVAAG 12

RESULT 15

US-08-480-070C-24
; Sequence 24, Application US/08480070C
; Patent No. 6066498
; GENERAL INFORMATION:
; APPLICANT: Levinson, Douglas A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF IMMUNE DISORDERS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,070C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-024
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 49 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-480-070C-24

Query Match 66.0%; Score 33; DB 3; Length 49;
Best Local Similarity 70.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VFAAGAAAAG 10
|||
Db 3 VFLAGGVAAG 12

Search completed: March 6, 2001, 12:53:53
Job time: 353 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:50:57 ; Search time 70.34 Seconds
(without alignments)
10.619 Million cell updates/sec

Title: us-09-196-161d-8

Perfect score: 50

Sequence: 1 VFAAGAAAAGV 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	100.0	395	2 A46031	immobilization sur
2	39	78.0	280	2 D72642	hypothetical prote
3	38	76.0	163	2 T45018	chemotaxis protein
4	38	76.0	641	2 G28771	hypothetical prote
5	38	76.0	3535	2 E83641	probable hemagglut
6	37	74.0	173	2 S18035	pathogenesis-relat
7	37	74.0	173	2 S18034	thumatin-like pro
8	37	74.0	173	2 S16524	hypothetical prote
9	37	74.0	360	2 G70615	probable amB prot
10	37	74.0	394	2 A70842	probable transket
11	37	74.0	630	2 B71276	hypothetical prote
12	36	72.0	299	2 T35765	hypothetical prote
13	36	72.0	363	2 T34931	pectin lyase (EC 4
14	36	72.0	379	2 S17979	hypothetical prote
15	36	72.0	441	2 H82642	4-hydroxyphenylace
16	36	72.0	480	1 C69378	hypothetical prote
17	36	72.0	680	2 T30620	hypothetical prote
18	36	72.0	3429	2 T13853	hypothetical prote
19	36	72.0	4845	2 T31067	BIR repeat contain
20	35	70.0	145	2 S35159	photosystem I chai
21	35	70.0	168	2 G70587	hypothetical prote
22	35	70.0	265	2 S19923	homeotic protein e
23	35	70.0	305	1 C45822	beta-lactamase (EC
24	35	70.0	312	2 A83486	cobalamin biosynth
25	35	70.0	393	2 T20268	hypothetical prote
26	35	70.0	394	2 A55045	probable 3-hydroxy
27	35	70.0	488	2 A82984	hypothetical prote
28	35	70.0	512	2 D82444	NAD(P) transhydrog
29	35	70.0	556	2 A70560	hypothetical prote

30 35 70.0 1629 2 T06461 DNA-binding protei
31 35 70.0 4613 2 T17409 polyketide synthas
32 34 68.0 118 2 F69498 hypothetical prote
33 34 68.0 130 2 A71530 ribosomal protein
34 34 68.0 130 2 H81684 ribosomal protein
35 34 68.0 171 2 S78525 alpha-amylose inhi
36 34 68.0 182 1 Y0KB11 type 1 fimbrial pr
37 34 68.0 211 2 T32591 hypothetical prote
38 34 68.0 233 2 D70058 hypothetical prote
39 34 68.0 349 2 C72630 probable threonine
40 34 68.0 365 2 T12040 cysteine proteinas
41 34 68.0 367 2 C70858 hypothetical prote
42 34 68.0 374 2 T03875 probable homeobox
43 34 68.0 375 2 T03874 probable homeobox
44 34 68.0 411 2 S41945 hypothetical prote
45 34 68.0 411 2 T36607 probable cell surf

ALIGNMENTS

RESULT 1

A46031

immobilization surface I-antigen precursor - Ichthyophthirius multifiliis (fragment)
C:Species: Ichthyophthirius multifiliis

C:Date: 29-Sep-1999 #sequence_revision 29-Sep-1999 #text_change 07-Dec-1999

C:Accession: A46031

R:Clark, T.G.; McGraw, R.A.; Dickerson, H.W.

Proc. Natl. Acad. Sci. U.S.A. 89, 6363-6367, 1992

A:Title: Developmental expression of surface antigen genes in the parasitic ciliate I

A:Reference number: A46031; MUID:92335298

A:Accession: A46031

A:Molecule type: mRNA; protein

A:Residues: 1-395 <CLA>

A:Cross-references: GB:M92907; MID:g3628568; PIDN:AAC36158.1; PID:g3628569

A:Note: the authors translated the codon UUG for residue 330 as lle

A:Note: sequence extracted from NCHI backbone (NCBIN:108734, NCBIP:108735); the sequ

C:Genetics:

A:Genetic code: SGC5

C:Keywords: glycoprotein; surface antigen

F:2-395/Product: immobilization surface I-antigen #status experimental <MAT>

F:156,191,245,281/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0% Score 50; DB 2; Length 395;
Best Local Similarity 100.0%; Pred. No. 0.85;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 VFAAGAAAAGV 11

Db 131 VFAAGAAAAGV 141

RESULT 2

D72642

hypothetical protein APE0572 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C:Accession: D72642

R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Ta

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.

DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero

A:Reference number: A72450; MUID:99310339

A:Accession: D72642

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-280 <KAW>

A:Cross-references: DDBJ:AP000060; NID:g5104188; PIDN:BAA79540.1; PID:g5104225

A:Experimental source: strain K1

C:Genetics:

A:Gene: APE0572

C:Superfamily: hypothetical protein MJ1440

Query Match 78.0%; Score 39; DB 2; Length 280;
Best Local Similarity 72.7%; Pred. No. 32;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VFAGAAAAGV 11
:: |||||
Db 233 LYAGAAAAGV 243

RESULT 3
T45018
chemotaxis protein, CheWII [imported] - Rhodobacter sphaeroides
C:Species: Rhodobacter sphaeroides
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 11-May-2000
C:Accession: T45018
R:Hamblin, P.A.; Maguire, B.A.; Grishanin, R.N.; Armitage, J.P.
Mol. Microbiol. 26, 1083-1096, 1997
A:Title: Evidence for dual chemosensory pathways in Rhodobacter sphaeroides.
A:Reference number: 222898; MUID:98086113
A:Accession: T45018
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-163 <HAM>
A:Cross-references: EMBL:AJ000977; PIDN:CAA04430.1
A:Experimental source: strain WS8
C:Genetics:
A:Gene: cheWII
C:Superfamily: chemotaxis cheW protein

Query Match 76.0%; Score 38; DB 2; Length 163;
Best Local Similarity 72.7%; Pred. No. 29;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VFAGAAAAGV 11
::||| |::|||
Db 150 IFAAQAAGV 160

RESULT 4
G28771
hypothetical protein C2814 (photosynthetic gene cluster) - Rhodobacter capsulatus
C:Species: Rhodobacter capsulatus
C:Date: 18-Oct-1989 #sequence_revision 18-Oct-1989 #text_change 04-Sep-1998
C:Accession: G28771
R:Yovan, D.C.; Bylina, E.J.; Alberti, M.; Begusch, H.; Hearst, J.E.
Cell 37, 949-957, 1984
A:Title: Nucleotide and deduced polypeptide sequences of the photosynthetic reaction-cent
A:Reference number: A90850; MUID:84259352
A:Accession: G28771
A:Molecule type: DNA
A:Residues: 1-641 <YOU>
A:Cross-references: GB:K01183
C:Superfamily: hypothetical protein C2814

Query Match 76.0%; Score 38; DB 2; Length 641;
Best Local Similarity 80.0%; Pred. No. 93;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 FAAGAAAAGV 11
|||| |||||
Db 378 FAAGMAAGL 387

RESULT 5
E83641
probable hemagglutinin PA0041 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Sep-2000
C:Accession: E83641

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950
A:Accession: E83641
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3535 <STO>
A:Cross-references: GB:AE004443; GB:AE004091; NID:g9945853; PIDN:AA03431.1; GSPDB:CN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0041

Query Match 76.0%; Score 38; DB 2; Length 3535;
Best Local Similarity 88.9%; Pred. No. 4e+02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FAAGAAAAG 10
|| |||||
Db 3129 FATGAAAAG 3137

RESULT 6
S18035
pathogenesis-related protein 1c precursor - barley
N:Alternate names: thaumatin-like protein 1c
C:Species: Hordeum vulgare (barley)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Aug-1999
C:Accession: S18035
R:Hahn, M.; Lehnackers, H.; Knogge, W.
submitted to the EMBL Data Library, March 1991
A:Reference number: S18033
A:Accession: S18035
A:Molecule type: mRNA
A:Residues: 1-173 <HAH>
A:Cross-references: EMBL:X58565; NID:g19084; PIDN:CAA41445.1; PID:g19085
C:Genetics:
A:Map position: 1
C:Superfamily: thaumatin I
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-173/Product: pathogenesis-related protein 1c #status predicted <MAT>

Query Match 74.0%; Score 37; DB 2; Length 173;
Best Local Similarity 88.9%; Pred. No. 44;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VFAAGAAA 9
|||||||
Db 13 VFAAGASAA 21

RESULT 7
S18034
pathogenesis-related protein 1 (a and b) precursor - barley
N:Alternate names: thaumatin-like protein 1a; thaumatin-like protein 1b
C:Species: Hordeum vulgare (barley)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Aug-1999
C:Accession: S18034; S70445
R:Hahn, M.; Lehnackers, H.; Knogge, W.
submitted to the EMBL Data Library, March 1991
A:Reference number: S18033
A:Accession: S18034
A:Molecule type: mRNA
A:Residues: 1-173 <HAH>
A:Cross-references: EMBL:X58566; NID:g19082; PIDN:CAA41446.1; PID:g19083
A:Note: pathogenesis-related protein 1b
A:Accession: S18033
A:Molecule type: mRNA
A:Residues: 1-173 <HA2>

A:Cross-references: EMBL:X58564; NID:g19080; PIDN:CAA1444.1; PID:g19081
 A:Note: pathogenesis-related protein 1a
 R:Bryngelsson, T.; Green, B.

Physiol. Mol. Plant Pathol. 35, 45-52, 1989

A:Title: Characterization of a pathogenesis-related, thaumatin-like protein isolated from

A:Reference number: S70445

A:Accession: S70445

A:Molecule type: protein

A:Residues: 21-46, X', 48

A:Note: accumulates in plants infected by viruses, bacteria or fungi

C:Genetics:

A:Map position: 1

C:Superfamily: thaumatin I

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-173/Product: pathogenesis-related protein 1 (a and b) #status predicted <MAT>

Query Match 74.0%; Score 37; DB 2; Length 173;

Best Local Similarity 88.9%; Pred. No. 44;

Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VFAGAAAA 9

Db 13 VFAGASAA 21

|||||:||

|||||:||

RESULT 8

S16524

thaumatin-like protein precursor - wheat

C:Species: Triticum aestivum (common wheat)

C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Aug-1999

C:Accession: S16524; S15461

R:Rebmann, G.; Mauch, F.; Dudler, R.

Plant Mol. Biol. 17, 283-285, 1991

A:Title: Sequence of a wheat cDNA encoding a pathogen-induced thaumatin-like protein.

A:Reference number: S16524; MUID:91322528

A:Accession: S16524

A:Molecule type: mRNA

A:Residues: 1-173 <REB>

A:Cross-references: EMBL:X58394; NID:g21878; PIDN:CAA41283.1; PID:g21879

A:Experimental source: strain c.v. Cheyenne

C:Superfamily: thaumatin I

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-173/Product: thaumatin-like protein #status predicted <MAT>

Query Match 74.0%; Score 37; DB 2; Length 173;

Best Local Similarity 88.9%; Pred. No. 44;

Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VFAGAAAA 9

Db 13 VFAGASAA 21

|||||:||

|||||:||

RESULT 9

G70615

hypothetical protein Rv0132c - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: G70615

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:96295987

A:Accession: G70615

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-360 <COL>

A:Cross-references: GB:292770; GB:AL123456; NID:g3261720; PIDN:CAB07038.1; PID:e306643;

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: Rv0132c

Query Match 74.0%; Score 37; DB 2; Length 360;

Best Local Similarity 88.9%; Pred. No. 82;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 FAAGAAAAG 10

Db 244 FAAGAAAG 252

||||| |||

||||| |||

RESULT 10

A70842

probable amB protein - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: A70842

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:98295987

A:Accession: A70842

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-394 <COL>

A:Cross-references: GB:AL021841; GB:AL123456; NID:g3261517; PIDN:CAA17078.1; PID:g289

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: amB

C:Superfamily: hippurate hydrolase

Query Match 74.0%; Score 37; DB 2; Length 394;

Best Local Similarity 88.9%; Pred. No. 88;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 FAAGAAAAG 10

Db 259 FAAGAAAG 267

||||| |||

||||| |||

RESULT 11

B71276

probable transketolase B (tktB) - syphilis spirochete

C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999

C:Accession: B71276

R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G

rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M

they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.

Science 281, 375-388, 1998

A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A:Reference number: A71250; MUID:98332770

A:Accession: B71276

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-630 <COL>

A:Cross-references: GB:AE001253; GB:AE000520; NID:g3323133; PIDN:AAC65792.1; PID:g332

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0824

C:Superfamily: hypothetical protein C2814

Query Match 74.0%; Score 37; DB 2; Length 630;

Best Local Similarity 80.0%; Pred. No. 1.3e+02;

Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 FAAGAAAAGV 11
 ||||| |||||
 Db 373 FAAGLACAGV 382

RESULT 12

T35765
 hypothetical protein SC7H2.34 - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
 C:Accession: T35765
 R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
 submitted to the EMBL Data Library, August 1999
 A:Reference number: Z21588
 A:Accession: T35765
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-299 <SAU>
 A:Cross-references: EMBL:AL109732; PIDN:CA52076.1; GSPDB:GN00070; SCOEDB:SC7H2.34
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOEDB:SC7H2.34

Query Match 72.0%; Score 36; DB 2; Length 299;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AAGAAAAG 10
 ||||| |||||
 Db 243 AAGAAAAG 250

RESULT 13

T34931
 hypothetical protein SC3F9.09 - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
 C:Accession: T34931
 R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, June 1998
 A:Reference number: Z21562
 A:Accession: T34931
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-363 <SEE>
 A:Cross-references: EMBL:AL023862; PIDN:CAAL19632.1; GSPDB:GN00070; SCOEDB:SC3F9.09
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOEDB:SC3F9.09

Query Match 72.0%; Score 36; DB 2; Length 363;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AGAAAAGV 11
 ||||| |||||
 Db 76 AGAAAAGV 83

RESULT 14

S17979
 pectin lyase (EC 4.2.2.10) A precursor - Aspergillus niger
 C:Species: Aspergillus niger
 C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
 C:Accession: S17979; S14116
 R:Kusters-van Someren, M.A.; Harmsen, J.A.M.; Kester, H.C.M.; Visser, J.
 Curr. Genet. 20, 293-299, 1991
 A:Title: Structure of the Aspergillus niger pectin lyase gene and its expression in Aspergillus
 A:Reference number: S17979; MUID:92035100
 A:Accession: S17979

A:Molecule type: DNA
 A:Residues: 1-379 <KUS>
 A:Cross-references: EMBL:X60724; NID:g2374; PIDN:CAA43130.1; PID:g2375
 R:Harmsen, J.A.M.; Kusters-van Someren, M.A.; Visser, J.
 Curr. Genet. 18, 161-166, 1990
 A:Title: Cloning and expression of a second Aspergillus niger pectin lyase gene (pela
 A:Reference number: S14116; MUID:91029617
 A:Accession: S14116
 A:Molecule type: DNA
 A:Residues: 1-49 <HAR>
 A:Cross-references: GB:X55784; NID:g288127; PIDN:CAA39305.1; PID:g288128
 C:Genetics:
 A:Gene: pela
 A:Introns: 116/2; 158/2; 190/3; 268/3
 C:Keywords: carbon-oxygen lyase
 F:1-20/Domain: signal sequence #status predicted <SIG>
 F:21-379/Product: pectin lyase A #status predicted <MAT>

Query Match 72.0%; Score 36; DB 2; Length 379;
 Best Local Similarity 72.7%; Pred. No. 1.2e+02;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 VFAAGAAAAGV 11
 ||||| |||||
 Db 13 VFAGSAAAAGV 23

RESULT 15

H82642
 hypothetical protein XF1738 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: H82642
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: H82642
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-441 <SIM>
 A:Cross-references: GB:AE003997; GB:AE003849; NID:g9106805; PIDN:AAF84547.1; GSPDB:GN
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.W.; Miracca, E.C.; Miyaki, C.
 , F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Miyaki, C.
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Silva
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
 M.; Tshuko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF1738

Query Match 72.0%; Score 36; DB 2; Length 441;
 Best Local Similarity 63.8%; Pred. No. 1.4e+02;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 VFAAGAAAAGV 11
 :||||| |:
 Db 160 LFAAGATAGGL 170

Fri Mar 9 15:48:07 2001

us-09-196-161d-8.rpr

Page 5

Search completed: March 6, 2001, 12:50:59
Job time: 180 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:54:41 ; Search time 38:83 Seconds
(without alignments)
9.148 Million cell updates/sec

Title: us-09-196-161d-8
Perfect score: 50
Sequence: 1 VFAAGAAAAGV 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	76.0	641	DXS_RHOCA	P26242 rhodobacter
2	37	74.0	169	RST1_AVESA	P30695 avena sativ
3	37	74.0	169	RST2_AVESA	P30696 avena sativ
4	37	74.0	169	RST3_AVESA	P30697 avena sativ
5	37	74.0	173	PR1A_HORVU	P32937 hordeum vul
6	37	74.0	173	PR1C_HORVU	P32938 hordeum vul
7	37	74.0	173	TLP_WHEAT	P27357 triticum ae
8	36	72.0	379	PLYA_ASPEG	Q01172 aspergillus
9	36	72.0	674	PTGA_BRELA	Q45298 brevibacter
10	36	72.0	1324	IRS2_HUMAN	Q9y4h2 homo sapien
11	35	70.0	145	PSAN_HORVU	P31093 hordeum vul
12	35	70.0	168	MB12_MYCTU	O05822 mycobacteri
13	35	70.0	265	YPR5_ECOLI	P22994 escherichia
14	35	70.0	305	BLAC_STRIA	P35393 streptomyce
15	35	70.0	312	PYDB_ENTFA	Q47741 enterococcu
16	35	70.0	389	ALR_MYCSM	P94967 mycobacteri
17	34	68.0	129	RL7_CHLMU	P38001 chlamydia m
18	34	68.0	129	RL7_CHLTR	O84318 chlamydia t
19	34	68.0	171	TAAD_HORVU	P11643 hordeum vul
20	34	68.0	182	FM12_KLEPN	P12903 klebsiella
21	34	68.0	236	LYSE_CORGL	P94633 corynebacte
22	34	68.0	367	TRMO_MYCTU	O53271 mycobacteri
23	34	68.0	411	YDNK_STRCO	P40180 streptomyce
24	34	68.0	428	FXB2_MOUSE	O64733 mus musculu
25	34	68.0	446	NRTA_SYNY3	P73452 synechocyst
26	34	68.0	643	MAOM_ASCSU	P27443 ascaris suu
27	34	68.0	1030	FBPI_DROME	Q04691 drosophila
28	34	68.0	1058	UL52_HSV11	P10236 herpes simp
29	33	66.0	106	RLA3_YEAST	P10622 saccharomyc
30	33	66.0	130	RL7_MYCTU	P37381 mycobacteri
31	33	66.0	194	YGG8_YEAST	P53163 saccharomyc
32	33	66.0	208	AC20_TENNO	P26967 tenebrio mo
33	33	66.0	312	SOX3_CHICK	P48433 gallus gall

RESULT 1

DXS_RHOCA
ID DXS_RHOCA STANDARD: PRT: 641 AA.
AC P26242;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PROBABLE 1-DEOXYXYLOSE-5-PHOSPHATE SYNTHASE (DXP SYNTHASE)
DE (PROTEIN C2814).
GN DXS.
OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84259352; PubMed=6744416;
RA Youvan D.C., Bylina E.J., Alberti M., Begusch H., Hearst J.E.;
RT "Nucleotide and deduced polypeptide sequences of the photosynthetic
RT reaction-center, B870 antenna, and flanking polypeptides from R.
RT capsulata.";
RL Cell 37:949-957(1984).
CC -!- FUNCTION: CATALYZES THE ACYLOIN CONDENSATION REACTION BETWEEN C
CC ATOMS 2 AND 3 OF PYRUVATE AND GLYCERALDEHYDE 3-PHOSPHATE TO YIELD
CC 1-DEOXY-D-XYLULOSE-5-PHOSPHATE (DXP) (BY SIMILARITY).
CC -!- COFACTOR: THIAMINE PYROPHOSPHATE (BY SIMILARITY).
CC -!- PATHWAY: IN THE BIOSYNTHETIC PATHWAY TO ISOPRENOLIDS, THIAMINE, AND
CC PYRIDOXOL (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE TRANSKETOLASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; K01183; -; NOT_ANNOTATED_CDS.
DR EMBL; Z11165; CAA77557.1; -;
DR PIR; G28771; G28771.
DR INTERPRO; IPR000360; -;
DR INTERPRO; IPR000399; -;
DR PROSITE; PS00801; TRANSKETOLASE_1; 1.
DR PROSITE; PS00802; TRANSKETOLASE_2; 1.
DR PROSITE; PS00187; TPP_ENZYMES; UNKNOWN_1.
KW Flavoprotein; Thiamine pyrophosphate; Isoprene biosynthesis;
KW Thiamine biosynthesis.
SQ SEQUENCE 641 AA; 67943 MW; CBCC0BF23C012201 CRC64;

Query Match 76.0%; Score 38; DB 1; Length 641;

Best Local Similarity 80.0%; Pred. No. 49;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 2 FAAGAAAAGV 11

|||||

```
Db 378 FAAGMAAGL 387
RESULT 2
RSTL_AVES
ID RSTL_AVES STANDARD; PRT; 169 AA.
AC P50695;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE THAUMATIN-LIKE PATHOGENESIS-RELATED PROTEIN 1 PRECURSOR.
GN RASTL-1.
OS Avena sativa (Oat).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Avena.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, RODNEY;
RX MEDLINE=96303536; PubMed=8755626;
RA Lin K.C., Bushnell W.R., Szabo L.J., Smith A.G.;
RT "Isolation and expression of a host response gene family encoding
thaumatin-like proteins in incompatible oat-stem rust fungus
interactions.";
RL Mol. Plant Microbe Interact. 9:511-522(1996).
CC -!- FUNCTION: ASSOCIATED WITH RESISTANCE AGAINST STEM RUST FUNGI.
CC -!- SIMILARITY: BELONGS TO THE THAUMATIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L39775; AAB09225.1; -
CC HSSP: P02883; ITHU.
CC INTERPRO: IPR001938; -
CC PFAM: PF00314; Thaumatin; 2.
CC PRINTS: PR00347; THAUMATIN.
CC PROSITE: PS00316; THAUMATIN.
CC Pathogenesis-related protein; Fungicide; Multigene family; Signal.
KW SIGNAL
FT CHAIN 1 21 POTENTIAL.
FT CHAIN 22 169 THAUMATIN-LIKE PATHOGENESIS-RELATED
FT PROTEIN 2.
SQ SEQUENCE 169 AA; 17403 MW; 0030CB363CCBE79 CRC64;

Query Match 74.0%; Score 37; DB 1; Length 169;
Best Local Similarity 88.9%; Pred. No. 24;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VFAAGAAA 9
Db 14 VFAAGASAA 22
|||||:|
RESULT 4
RST3_AVES
ID RST3_AVES STANDARD; PRT; 169 AA.
AC P50697;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE THAUMATIN-LIKE PATHOGENESIS-RELATED PROTEIN 3 PRECURSOR.
GN RASTL-3.
OS Avena sativa (Oat).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Avena.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, RODNEY;
RX MEDLINE=96303536; PubMed=8755626;
RA Lin K.C., Bushnell W.R., Szabo L.J., Smith A.G.;
RT "Isolation and expression of a host response gene family encoding
thaumatin-like proteins in incompatible oat-stem rust fungus
interactions.";
RL Mol. Plant Microbe Interact. 9:511-522(1996).
CC -!- FUNCTION: ASSOCIATED WITH RESISTANCE AGAINST STEM RUST FUNGI.
CC -!- SIMILARITY: BELONGS TO THE THAUMATIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L39776; AAB09226.1; -
CC HSSP: P02883; ITHU.
CC INTERPRO: IPR001938; -

Query Match 74.0%; Score 37; DB 1; Length 169;
Best Local Similarity 88.9%; Pred. No. 24;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VFAAGAAA 9
Db 14 VFAAGASAA 22
|||||:|
RESULT 3
RST2_AVES
ID RST2_AVES STANDARD; PRT; 169 AA.
AC P50696;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE THAUMATIN-LIKE PATHOGENESIS-RELATED PROTEIN 2 PRECURSOR.
GN RASTL-2.
OS Avena sativa (Oat).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Avena.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, RODNEY;
RX MEDLINE=96303536; PubMed=8755626;
```

```

DR PFAM: PF00314; thaumatin; 2.
DR PRINTS: PR00347; THAUMATIN.
DR PROSITE: PS00316; THAUMATIN; 1.
KW Pathogenesis-related protein; Fungicide; Multigene family; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 169 THAUMATIN-LIKE PATHOGENESIS-RELATED
          PROTEIN 3.
SQ SEQUENCE 169 AA; 17328 MW; 647CA7D6D54108AF CRC64;

Query Match          74.0%; Score 37; DB 1; Length 169;
Best Local Similarity 88.9%; Pred. No. 24;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VFAAGAAAA 9
   |||||:|
Db 14 VFAAGASAA 22

RESULT 5
PR1A_HORVU
ID PR1A_HORVU STANDARD; PRT; 173 AA.
AC P32937;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE PATHOGENESIS-RELATED PROTEIN 1A/1B PRECURSOR.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Poaceae; Hordeum.
OC Magnoliophyta; Liliopsida; Poales;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=LEAF;
RA Hahn M., Lehnackers H., Knogge W.;
RT "Cultivar-specific elicitation of barley defense reactions by the
   phytoxic peptide NIPI from Rhynchosporium secalis.";
RL Mol. Plant Microbe Interact. 6:745-754(1993).
CC -!- SIMILARITY: BELONGS TO THE THAUMATIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X58564; CAA41444.1; -
DR PIR: S18033; S18033.
DR HSP: P02883; 1THU.
DR INTERPRO: IPR001938; -
DR PFAM: PF00314; thaumatin; 2.
DR PRINTS: PR00347; THAUMATIN.
DR PROSITE: PS00316; THAUMATIN; 1.
KW Pathogenesis-related protein; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 173 PATHOGENESIS-RELATED PROTEIN 1A/1B.
SQ SEQUENCE 173 AA; 17548 MW; 9BBFA89E882802BD CRC64;

Query Match          74.0%; Score 37; DB 1; Length 173;
Best Local Similarity 88.9%; Pred. No. 25;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VFAAGAAAA 9
   |||||:|
Db 13 VFAAGASAA 21

RESULT 6
PR1A_HORVU
ID PR1A_HORVU STANDARD; PRT; 173 AA.
AC P32937;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE PATHOGENESIS-RELATED PROTEIN 1A/1B PRECURSOR.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Poaceae; Hordeum.
OC Magnoliophyta; Liliopsida; Poales;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=LEAF;
RA Hahn M., Lehnackers H., Knogge W.;
RT "Cultivar-specific elicitation of barley defense reactions by the
   phytoxic peptide NIPI from Rhynchosporium secalis.";
RL Mol. Plant Microbe Interact. 6:745-754(1993).
CC -!- SIMILARITY: BELONGS TO THE THAUMATIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X58564; CAA41444.1; -
DR PIR: S18033; S18033.
DR HSP: P02883; 1THU.
DR INTERPRO: IPR001938; -
DR PFAM: PF00314; thaumatin; 2.
DR PRINTS: PR00347; THAUMATIN.
DR PROSITE: PS00316; THAUMATIN; 1.
KW Pathogenesis-related protein; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 173 PATHOGENESIS-RELATED PROTEIN 1A/1B.
SQ SEQUENCE 173 AA; 17548 MW; 9BBFA89E882802BD CRC64;

Query Match          74.0%; Score 37; DB 1; Length 173;
Best Local Similarity 88.9%; Pred. No. 25;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VFAAGAAAA 9
   |||||:|
Db 13 VFAAGASAA 21

RESULT 6
PR1A_HORVU
ID PR1A_HORVU STANDARD; PRT; 173 AA.
AC P32937;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE PATHOGENESIS-RELATED PROTEIN 1C PRECURSOR.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Poaceae; Hordeum.
OC Magnoliophyta; Liliopsida; Poales;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=LEAF;
RA Hahn M., Lehnackers H., Knogge W.;
RT "Cultivar-specific elicitation of barley defense reactions by the
   phytoxic peptide NIPI from Rhynchosporium secalis.";
RL Mol. Plant Microbe Interact. 6:745-754(1993).
CC -!- SIMILARITY: BELONGS TO THE THAUMATIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X58565; CAA41445.1; -
DR PIR: S18035; S18035.
DR HSP: P02883; 1THU.
DR INTERPRO: IPR001938; -
DR PFAM: PF00314; thaumatin; 2.
DR PRINTS: PR00347; THAUMATIN.
DR PROSITE: PS00316; THAUMATIN; 1.
KW Pathogenesis-related protein; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 173 PATHOGENESIS-RELATED PROTEIN 1C.
SQ SEQUENCE 173 AA; 17576 MW; 99D0550E882BCE72 CRC64;

Query Match          74.0%; Score 37; DB 1; Length 173;
Best Local Similarity 88.9%; Pred. No. 25;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VFAAGAAAA 9
   |||||:|
Db 13 VFAAGASAA 21

RESULT 7
TLIP_WHEAT
ID TLIP_WHEAT STANDARD; PRT; 173 AA.
AC P27357;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DE THAUMATIN-LIKE PROTEIN PWIR2 PRECURSOR.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Poaceae; Triticum.
OC Magnoliophyta; Liliopsida; Poales;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CV CHEYENNE; TISSUE=LEAF;
RA MEDLINE=91322526; PubMed=1863782;
RA Reimann G., Mauch F., Dudler R.;
RT "Sequence of a wheat cDNA encoding a pathogen-induced thaumatin-like
   protein.";
RL Plant Mol. Biol. 17:283-285(1991).
CC -!- INDUCTION: BY PATHOGENS.
CC -!- SIMILARITY: BELONGS TO THE THAUMATIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X58394; CAA41283.1; -
 CC PIR; S15461; S15461.
 CC PIR; S16524; S16524.
 CC HSSP; P02883; LTHU.
 CC INTERPRO; IPR001938; -
 CC PFAM; PF00314; Thaumatin; 2.
 CC PRINTS; PR00347; THAUMATIN.
 CC PROSITE; PS00316; THAUMATIN; 1.
 CC Pathogenesis-related protein; Signal.
 CC FT SIGNAL 1 20 POTENTIAL.
 CC FT CHAIN 21 173 THAUMATIN-LIKE PROTEIN PWIR2.
 CC SEQUENCE 173 AA; 17604 MW; 3704BDBDBES1365C CRC64;
 CC -----
 CC Query Match 74.0%; Score 37; DB 1; Length 173;
 CC Best Local Similarity 88.9%; Pred. No. 25;
 CC Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 CC -----
 CC QY 1 VFAAGAAAA 9
 CC IIIII:II
 CC Db 13 VFAAGASAA 21
 CC -----
 CC RESULT 8
 CC PLVA_ASPNG STANDARD; PRT; 379 AA.
 CC AC Q01172;
 CC DT 01-JUL-1993 (Rel. 26, Created)
 CC DT 01-JUL-1993 (Rel. 26, Last sequence update)
 CC DT 15-DEC-1998 (Rel. 37, Last annotation update)
 CC DE PECTIN LYASE A PRECURSOR (EC 4.2.2.10) (PLA) (PECTIN LYASE II) (PLII).
 CC GN PELA.
 CC OS Aspergillus niger.
 CC OC Eukaryota; Fungi; Ascomycota; Eurotiales; Trichocomaceae;
 CC OC anamorphic Trichocomaceae; Aspergillus.
 CC RN [1]
 CC RP SEQUENCE FROM N.A., AND SEQUENCE OF 21-40 AND 115-125.
 CC RC STRAIN=CBS 120.49 / N400;
 CC RX MEDLINE=92035100; PubMed=1934134;
 CC RA Kusters-Van Someren M.A., Harmsen J.A.M., Kester H.C.M., Visser J.;
 CC RT "Structure of the Aspergillus niger pella gene and its expression in
 CC RT Aspergillus niger and Aspergillus nidulans."
 CC RL Curr. Genet. 20:293-299(1991).
 CC RN [2]
 CC RP SEQUENCE OF 1-49 FROM N.A.
 CC RC STRAIN=CBS 120.49 / N400;
 CC RX MEDLINE=91029617; PubMed=2225145;
 CC RA Harmsen J.A.M., Kuster-Van Someren M.A., Visser J.;
 CC RT "Cloning and expression of a second Aspergillus niger pectin lyase
 CC RT gene (pela): indications of a pectin lyase gene family in A. niger."
 CC RL Curr. Genet. 18:161-166(1990).
 CC RN [3]
 CC RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
 CC RC STRAIN=CBS 120.49 / N400;
 CC RX MEDLINE=97341230; PubMed=9195887;
 CC RA Mayans O., Scott M., Connerton I., Gravesen T., Benen J., Visser J.,
 CC RA Pickersgill R., Jenkins J.;
 CC RT "Two crystal structures of pectin lyase A from Aspergillus reveal a
 CC RT pH driven conformational change and striking divergence in the
 CC RT substrate-binding clefts of pectin and pectate lyases."
 CC RL Structure 5:677-689(1997).
 CC CC -!- CATALYTIC ACTIVITY: ELIMINATIVE CLEAVAGE OF PECTIN TO GIVE
 CC CC OLIGOSACCHARIDES WITH TERMINAL 4-DEOXY-6-METHYL-ALPHA-D-GALACT-
 CC CC 4-ENURONOSYL GROUPS.
 CC CC -!- PTM: N-ACETYLGLUCOSAMINE WAS NOT FOUND IN PURIFIED PECTIN LYASE,
 CC BUT SOME O-GLYCOSYLATION MAY OCCUR.

CC -!- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X60724; CAA43130.1; -
 CC EMBL; X55784; CAA39305.1; -
 CC PIR; S17979; S17979.
 CC PDB; 1IDJ; 15-OCT-97.
 CC INTERPRO; IPR002022; -
 CC PFAM; PF00544; pec_lyase; 1.
 CC Lyase; Signal; Glycoprotein; Multigene family; 3D-structure.
 CC FT SIGNAL 1 20
 CC FT CHAIN 21 379 PECTIN LYASE A.
 CC FT ACT_SITE 256 256 POTENTIAL.
 CC FT CARBOHYD 88 88
 CC FT CARBOHYD 129 129 N-LINKED (GLCNAC...).
 CC SEQUENCE 379 AA; 39854 MW; 4DF1326BFFCA77B5 CRC64;
 CC -----
 CC Query Match 72.0%; Score 36; DB 1; Length 379;
 CC Best Local Similarity 72.7%; Pred. No. 65;
 CC Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 CC -----
 CC QY 1 VFAAGAAAAGV 11
 CC III IIII
 CC Db 13 VFAGSAAGV 23
 CC -----
 CC RESULT 9
 CC PTGA_BRELA STANDARD; PRT; 674 AA.
 CC AC Q45298;
 CC DT 01-NOV-1997 (Rel. 35, Created)
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
 CC DT 15-JUL-1999 (Rel. 38, Last annotation update)
 CC DE PTS SYSTEM, GLUCOSE-SPECIFIC IIABC COMPONENT (EIIABC-GLC) (GLUCOSE-
 CC DE PERMEASE IIABC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, ABC
 CC COMPONENT) (EC 2.7.1.69) (EII-GLC/EIIB-GLC).
 CC GN PTSG.
 CC OS Brevibacterium lactofermentum.
 CC OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC OC Actinomycetales; Corynebacterineae.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=ATCC 13869;
 CC RA Yoon K.H.;
 CC RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
 CC CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
 CC CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
 CC CC -TRANSPORT SYSTEM. THE IICD DOMAIN CONTAINS THE SUGAR BINDING SITE
 CC CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
 CC CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
 CC CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
 CC CC THE SUGAR.
 CC CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =
 CC CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
 CC CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC CC -!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.
 CC CC -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
 CC CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

or send an email to license@isb-sib.ch.

```

CC EMBL; L18975; AAA22992.1;
CC HSSP; P08837; 2F3G.
DR INTERPRO; IPR001127;
DR INTERPRO; IPR001996;
DR PFAM; PF00358; PTS_EIIA_1; 1.
DR PFAM; PF00367; PTS_EIIB; 1.
DR PROSITE; PS00371; PTS_EIIB_CVS; 1.
DR PROSITE; PS01035; PTS_EIIB_CVS; 1.
KW phosphotransferase system; Sugar transport; Transferase;
KW phosphorylation; Transmembrane.
FT DOMAIN 1 43 EIIA DOMAIN.
FT DOMAIN 542 674 EIIA DOMAIN.
FT TRANSMEM 126 146 POTENTIAL.
FT TRANSMEM 162 182 POTENTIAL.
FT TRANSMEM 193 213 POTENTIAL.
FT TRANSMEM 225 245 POTENTIAL.
FT TRANSMEM 260 280 POTENTIAL.
FT TRANSMEM 303 323 POTENTIAL.
FT TRANSMEM 344 364 POTENTIAL.
FT TRANSMEM 377 397 POTENTIAL.
FT TRANSMEM 409 429 POTENTIAL.
FT TRANSMEM 442 462 POTENTIAL.
FT TRANSMEM 493 513 POTENTIAL.
FT MOD_RES 28 28 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 594 594 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 674 AA; 71626 MW; 67A75AAF76E42FA2 CRC64;

Query Match 72.0%; Score 36; DB 1; Length 674;
Best Local Similarity 100.0%; Pred. No. 1e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 AAGAAAG 10
Db 497 AAGAAAG 504

RESULT 10
IRS2_HUMAN
ID IRS2_HUMAN STANDARD; PRT; 1324 AA.
AC Q9Y4H2;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE INSULIN RECEPTOR SUBSTRATE-2 (IRS-2).
GN IRS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-97460123; PubMed-9312143;
RA Ogihara T., Isobe T., Ichimura T., Taoka M., Funaki M., Sakoda H.,
RA Onishi Y., Inukai K., Anai M., Fukushima Y., Kikuchi M., Yazaki Y.,
RA Oka Y., Asano T.;
RT "14-3-3 protein binds to insulin receptor substrate-1, one of the
RT binding sites of which is in the phosphotyrosine binding domain.";
RL J. Biol. Chem. 272:25267-25274(1997).
CC -!- FUNCTION: MAY MEDIATE THE CONTROL OF VARIOUS CELLULAR PROCESSES
CC BY INSULIN.
CC -!- SIMILARITY: CONTAINS 1 PTF DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PTF DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

CC EMBL; AB000732; BAA24500.1;
CC MIM; 600797;
DR INTERPRO; IPR001849;
DR INTERPRO; IPR002404;
DR PFAM; PF00169; PH; 1.
DR PROSITE; PR00628; INSULINRS1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
KW Phosphorylation.
FT DOMAIN 16 144 PH.
FT DOMAIN 190 303 PTB.
FT MOD_RES 540 540 PHOSPHORYLATION (BY INSR) (BY
FT MOD_RES 653 653 SIMILARITY).
FT MOD_RES 675 675 PHOSPHORYLATION (BY INSR) (BY
FT MOD_RES 919 919 SIMILARITY).
FT MOD_RES 978 978 PHOSPHORYLATION (BY INSR) (BY
FT MOD_RES 1253 1253 SIMILARITY).
FT DOMAIN 19 28 POLY-ASN.
FT DOMAIN 371 380 POLY-ALA.
FT DOMAIN 447 452 POLY-SER.
FT DOMAIN 460 467 POLY-PRO.
FT DOMAIN 533 537 POLY-GLY.
FT DOMAIN 642 645 POLY-SER.
FT DOMAIN 694 701 POLY-ALA.
FT DOMAIN 944 947 POLY-SER.
FT DOMAIN 1031 1038 POLY-PRO.
FT DOMAIN 1265 1278 POLY-PRO.
SQ SEQUENCE 1324 AA; 136482 MW; 3D7B4AB2AE45104 CRC64;

Query Match 72.0%; Score 36; DB 1; Length 1324;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 AAGAAAG 10
Db 372 AAGAAAG 379

RESULT 11
PSAN_HORVU
ID PSAN_HORVU STANDARD; PRT; 145 AA.
AC P31093;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PHOTOSYSTEM I REACTION CENTRE SUBUNIT N PRECURSOR (PSI-N).
GN PSAN.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. SVALOFS BONUS; TISSUE-LEAF;
RX MEDLINE-95283637; PubMed-8507834;
RA Knoetzel J., Simpson D.J.;
RT "The primary structure of a cDNA for psan, encoding an extrinsic
RT luminal polypeptide of barley photosystem I.";
RL Plant Mol. Biol. 22:337-345(1993).
CC -!- FUNCTION: MAY FUNCTION IN MEDIATING THE BINDING OF THE ANTENNA
CC COMPLEXES TO THE PSI REACTION CENTRE AND CORE ANTENNA.
CC -!- SUBCELLULAR LOCATION: ON THE LUMENAL SURFACE OF THE THYLAKOID
CC MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE PSAN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X66428; CAA47056.1; -
 DR PIR; S24938; S24938.
 DR PIR; S35159; S35159.
 DR MENDEL; 8667; HORVU; PSAN1.
 KW Photosystem I: Photosynthesis; Chloroplast; Transit peptide;
 FT TRANSIT 1 60 CHLOROPLAST (BY SIMILARITY).
 FT CHAIN 61 145 PHOTOSYSTEM I REACTION CENTRE SUBUNIT N.
 SQ SEQUENCE 145 AA; 15500 MW; E99B6B6C5F984BCC CRC64;

Query Match 70.0%; Score 35; DB 1; Length 145;
 Best Local Similarity 80.0%; Pred. No. 43;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VFAAGAAAG 10
 |||||
 DB 47 VFAATAASAG 56

RESULT 12
 MB12_MYCTU STANDARD; PRT; 168 AA.
 AC 005822;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE LOW MOLECULAR WEIGHT ANTIGEN MTB12 PRECURSOR (LOW MOLECULAR WEIGHT
 DE PROTEIN ANTIGEN 2) (CPP-2).
 GN MTB12 OR CFP2 OR RV2376C OR MTCY27.04.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 49-77.
 RC STRAIN=H37RV;
 RX MEDLINE=98380366; PubMed=9712769;
 RA Webb J.R., Vedvick T.S., Alderson M.R., Guderian J.A., Jen S.S.,
 RA O'vendale P.J., Johnson S.M., Read S.G., Skeiky Y.A.W.;
 RT "Molecular cloning, expression, and immunogenicity of MTB12, a novel
 RT low-molecular-weight antigen secreted by Mycobacterium tuberculosis.";
 RL Infect. Immun. 66:4208-4214(1998).
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [3]
 RP SEQUENCE OF 49-63.

RC STRAIN=H37RV;
 RA Bhaskar S., Mukherjee R.;
 RT "Isolation, purification and immunological characterization of low
 RT molecular weight protein antigens from culture filtrate of
 RT M. tuberculosis.";
 RL Submitted (NOV-1998) to the SWISS-PROT data bank.
 CC -!- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF PROTECTIVE IMMUNE
 CC RESPONSES.

CC -!- SUBCELLULAR LOCATION: SECRETED.

CC -!- SIMILARITY: BELONGS TO THE MTB12 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: Z95208; CAB08476.1; -
 DR EMBL; AF062036; AAC35793.1; -
 DR TUBERCULIST; RV2376C; -
 KW Antigen; Signal.
 FT SIGNAL 1 22 POTENTIAL.
 FT PROPEP 23 48
 FT CHAIN 49 168 LOW MOLECULAR WEIGHT ANTIGEN MTB12.
 SQ SEQUENCE 168 AA; 16635 MW; A809937458009BFD CRC64;

Query Match 70.0%; Score 35; DB 1; Length 168;
 Best Local Similarity 88.9%; Pred. No. 48;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAGAAAAGV 11
 |
 DB 16 AIGAAAAAGV 24

RESULT 13
 YPR5_ECOLI STANDARD; PRT; 265 AA.
 AC P22994;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 01-AUG-1992 (Rel. 23, Last annotation update)
 DE HYPOTHETICAL 28.3 KDA PROTEIN IN PAR LOCUS (ORF 5).
 OS Escherichia coli.
 OG Plasmid Incp-beta RP4.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91035226; PubMed=2172207;
 RA Gerlitz M., Hrabak O., Schwab H.;
 RT "Partitioning of broad-host-range plasmid RP4 is a complex system
 RT involving site-specific recombination.";
 RL J. Bacteriol. 172:6194-6203(1990).
 CC -!- SUBCELLULAR LOCATION: SECRETED.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: M59825; AAA26413.1; -
 DR Plasmid; Hypothetical protein.
 SQ SEQUENCE 265 AA; 28300 MW; 0143EE20598DB717 CRC64;

Query Match 70.0%; Score 35; DB 1; Length 265;
 Best Local Similarity 70.0%; Pred. No. 70;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 FAAGAAAAGV 11
 |
 DB 57 FAAGVAAGGL 66

RESULT 14

RESULT 15	PYDB_ENTFA	STANDARD;	PRT;	312 AA.
ID	PYDB_ENTFA			
AC	Q47741;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	DIHYDROOrotate dehydrogenase B, catalytic subunit (EC 1.3.3.1)			
DE	(DIHYDROOrotate oxidase B) (DHODHase B) (DHODase B) (DHOD B).			
GN	PYDB or PYRD.			
OS	Enterococcus faecalis (Streptococcus faecalis).			
OC	Bacteria: Firmicutes; Bacillus/Clostridium group; Enterococcaceae;			
OC	Enterococcus.			
OC	[1]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=ATCC 47077 / OGI:RF;			
RC	MEDLINE=96074317; PubMed=7592480;			
RX				

Search completed: March 6, 2001, 12:54:43
Job time: 400 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:52:59 ; Search time 116.78 Seconds
(without alignments)
11.040 Million cell updates/sec

Title: US-09-196-161d-8
Perfect score: 50
Sequence: 1 VFAGAAAGV 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SP TREMBL:15:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	50	100.0	394	5 Q27208	Q27208 ichtyophth
2	50	100.0	442	5 Q9XZG2	Q9XZG2 ichtyophth
3	39	78.0	280	1 Q9YK66	Q9YK66 aeropyrum p
4	38	76.0	163	2 Q33555	Q33555 rhodobacter
5	37	74.0	173	10 Q9ZSN0	Q9ZSN0 secale cere
6	37	74.0	173	10 Q9S776	Q9S776 secale cere
7	37	74.0	360	2 P96809	P96809 mycobacteri
8	37	74.0	381	10 Q9LD54	Q9LD54 oryza sativ
9	37	74.0	394	2 Q53358	Q53358 mycobacteri
10	37	74.0	417	2 Q9RKH6	Q9RKH6 streptomyce
11	37	74.0	630	2 Q83796	Q83796 streponema p
12	37	74.0	1611	2 Q33957	Q33957 streptomyce
13	36	72.0	299	2 Q9S213	Q9S213 streptomyce
14	36	72.0	363	2 Q69949	Q69949 streptomyce
15	36	72.0	441	2 Q9PCP0	Q9PCP0 xylella fas
16	36	72.0	480	1 Q29235	Q29235 archaeoglob
17	36	72.0	530	10 Q9LJ18	Q9LJ18 oryza sativ
18	36	72.0	543	4 Q9UB19	Q9UB19 homo sapien
19	36	72.0	622	12 Q9W9L2	Q9W9L2 porcine ade

20	36	72.0	633	5 Q9W3G0	Q9W3G0 drosophila
21	36	72.0	640	12 Q11308	Q11308 molluscum c
22	36	72.0	680	12 Q98187	Q98187 molluscum c
23	36	72.0	735	4 Q9P227	Q9P227 homo sapien
24	36	72.0	903	5 Q9VKK0	Q9VKK0 drosophila
25	36	72.0	3429	5 Q24593	Q24593 drosophila
26	36	72.0	3429	5 Q9W425	Q9W425 drosophila
27	36	72.0	4829	4 Q9NR09	Q9NR09 homo sapien
28	36	72.0	4845	11 Q88738	Q88738 mus musculu
29	35	70.0	81	10 Q9M4Y6	Q9M4Y6 oryza sativ
30	35	70.0	195	5 Q9W2P9	Q9W2P9 drosophila
31	35	70.0	209	12 Q9QSR5	Q9QSR5 human immun
32	35	70.0	227	5 Q9VJQ9	Q9VJQ9 drosophila
33	35	70.0	265	5 Q26472	Q26472 schistocerc
34	35	70.0	349	2 P70903	P70903 borrelia he
35	35	70.0	370	5 Q9VTX3	Q9VTX3 drosophila
36	35	70.0	393	5 Q18880	Q18880 caenorhabdi
37	35	70.0	394	2 Q53924	Q53924 streptomyce
38	35	70.0	418	10 Q9LW14	Q9LW14 oryza sativ
39	35	70.0	479	10 Q9LWH4	Q9LWH4 oryza sativ
40	35	70.0	494	5 Q9N983	Q9N983 leishmania
41	35	70.0	512	2 Q9KM26	Q9KM26 vibrio chol
42	35	70.0	556	2 Q06152	Q06152 mycobacteri
43	35	70.0	715	5 Q9NKA6	Q9NKA6 drosophila
44	35	70.0	898	10 Q9SNE4	Q9SNE4 oryza sativ
45	35	70.0	1322	11 Q9QZP6	Q9QZP6 mus musculu

ALIGNMENTS

RESULT 1

Q27208
ID Q27208 PRELIMINARY; PRT; 394 AA.
AC Q27208;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE IMMOBILIZATION ANTIGEN PRECURSOR (FRAGMENT).
OS Ichthyophthirius multifiliis.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Ophryoglenina; Ichthyophthirius.
OX NCBI_TaxID=5932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GEORGIA;
RX MEDLINE=92335298; PubMed=1631132;
RA Clark T.G., McGraw R.A., Dickerson H.W.;
RT "Developmental expression of surface antigen genes in the parasitic ciliate Ichthyophthirius multifiliis.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:6363-6367(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GEORGIA;
RX MEDLINE=93020590; PubMed=1383510;
RA Lin T.L., Dickerson H.W.;
RT "Purification and partial characterization of immobilization antigens from Ichthyophthirius multifiliis.";
RL J. Protozool. 39:457-463(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=GEORGIA;
RA Clark T.;
RT Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=GEORGIA;
RA Clark T.;
RT Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: M92907; AAC36158.1; -
KW Signal.
FT NON_TER 1 1
FT SIGNAL <1 1 POTENTIAL.

SQ SEQUENCE 394 AA; 39495 MW; 3013C2B2BEFDB682 CRC64;

Query Match 100.0%; Score 50; DB 5; Length 394;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VFAAGAAAAGV 11
:|||||
Db 130 VFAAGAAAAGV 140

RESULT 2

ID Q9XZG2 PRELIMINARY; PRT; 442 AA.

AC Q9XZG2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE IMMOBILIZATION ANTIGEN PRECURSOR.
GN IAG48.
OS Ichthyophthirius multifiliis.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Ophryoglenina; Ichthyophthirius.
OX NCBI_TaxID=5932;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=GI;

RX MEDLINE=99196987; PubMed=10095108;
RA Clark T.G., Lin T.L., Jackwood D.A., Sherrill J., Lin Y.,

RA Dickinson H.W.;
RT "The gene for an abundant parasite coat protein predicts tandemly

RT repetitive metal binding domains.";
RL Gene 229:91-100(1999).
RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=GI;

RA Gaertig J., Gao Y., Tishgarten T., Clark T.G., Dickerson H.W.;
RT "Surface display of a parasite antigen in the ciliate Tetrahymena

RT thermophila.";
RL Nat. Biotechnol. 0:0-0(1999).

DR ENBL: AFI40273; AAD31283.1; -.

FT SIGNAL

FT CHAIN 1 20 POTENTIAL.

FT CHAIN 21 442 IMMOBILIZATION ANTIGEN.

SQ SEQUENCE 442 AA; 45025 MW; 52658F3F65D27AFA CRC64;

Query Match 100.0%; Score 50; DB 5; Length 442;

Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VFAAGAAAAGV 11
:|||||
Db 150 VFAAGAAAAGV 160

RESULT 3

ID Q9YEK6 PRELIMINARY; PRT; 280 AA.

AC Q9YEK6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE HYPOTHETICAL 28.6 KDA PROTEIN AFE0572.
GN AFE0572.

OS Aeropyrum pernix.

OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;

OC Aeropyrum.

OX NCBI_TaxID=56636;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=K1;

RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K., Kubota K., Nakamura Y.,
RA Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR ENBL: AP000060; BAA79540.1; -.
KW Hypothetical protein.
SQ SEQUENCE 280 AA; 28561 MW; B19A7E01C8CECCE7 CRC64;

Query Match 78.0%; Score 39; DB 1; Length 280;

Best Local Similarity 72.7%; Pred. No. 70;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VFAAGAAAAGV 11
:|||||
Db 233 LYFAGAAAAGV 243

RESULT 4

ID Q33555 PRELIMINARY; PRT; 163 AA.

AC Q33555;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE CHEMOTAXIS PROTEIN, CHEWII.
GN CHEWII.

OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).

OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;

OC Rhodobacter.

OX NCBI_TaxID=1063;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=WS8;

RX MEDLINE=98086113; PubMed=9426144;

RA Hamblin P.A., Maguire B.A., Grishanin R.N., Armitage J.P.;
RT "Evidence for two chemosensory pathways in Rhodobacter sphaeroides.";

RL Mol. Microbiol. 26:1083-1096(1997).

DR ENBL: AJ000977; CAA04430.1; -.

DR INTERPRO: IPR002545; -.

DR PFAM: PF01584; Chew; 1.

SQ SEQUENCE 163 AA; 17269 MW; E2990B0D9C3B4AB4 CRC64;

Query Match 76.0%; Score 38; DB 2; Length 163;

Best Local Similarity 72.7%; Pred. No. 60;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VFAAGAAAAGV 11
:|||||
Db 150 IFAAQASAGV 160

RESULT 5

ID Q9ZSNO PRELIMINARY; PRT; 173 AA.

AC Q9ZSNO;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE THAUMATIN-LIKE PROTEIN 4 PRECURSOR.
GN TLP4.

OS Secale cereale (Rye).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; Liliopsida; Poales; Poaceae; Secale.

OX NCBI_TaxID=4550;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-LEAF;
RA Chan Y.-W., Griffith M., Chow K.-C.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF100142; AAC83824.1; -;
DR HSSP; P02883; 1THV.
DR MENDEL; 36479; Secce:2330:36479.
DR INTERPRO; IPR001938; -;
DR PFAM; PF00314; thaumat. 2.
DR PRINTS; PR00347; THAUMATIN.
DR PROSITE; PS00316; THAUMATIN; UNKNOWN_1.
DR PRODOM; PD001321; -; 1.
SQ SEQUENCE 173 AA; 17618 MW; A992E460A5A30E4E CRC64;

Query Match 74.0%; Score 37; DB 10; Length 173;
Best Local Similarity 88.9%; Pred. No. 90;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VFAAGAAAA 9
DB 13 VFAAGASAA 21
|||||:|

RESULT 6

Q9S776 PRELIMINARY; PRT; 173 AA.
AC Q9S776;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE THAUMATIN-LIKE PROTEIN 2 PRECURSOR.
GN TLP2 OR TLP1 OR TLP3.
OS Secale cereale (Rye).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Secale.
OX NCBI_TaxID=4550;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LEAF;
RA Chan Y.-W., Griffith M., Chow K.-C.;
RT "Cloning of Pathogenesis-Related Type 5 Genes in Winter Rye."
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF099671; AAC83830.1; -;
DR EMBL; AF096927; AAC67259.1; -;
DR EMBL; AF099670; AAC83829.1; -;
DR HSSP; P02883; 1THV.
DR INTERPRO; IPR001938; -;
DR PFAM; PF00314; thaumat. 2.
DR PRINTS; PR00347; THAUMATIN.
DR PROSITE; PS00316; THAUMATIN; 1.
SQ SEQUENCE 173 AA; 17638 MW; C420001C88DFAE79 CRC64;

Query Match 74.0%; Score 37; DB 10; Length 173;
Best Local Similarity 88.9%; Pred. No. 90;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VFAAGAAAA 9
DB 13 VFAAGASAA 21
|||||:|

RESULT 7

P96809 PRELIMINARY; PRT; 360 AA.
AC P96809;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE HYPOTHETICAL 38.4 KDA PROTEIN.
GN RV0132C OR WTC15.06C.
OS Mycobacterium tuberculosis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN NCBI_TaxID=1773;
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL; Z92770; CAB07038.1; -;
DR TUBERCULIST; RV0132C; -;
KW Hypothetical protein.
SQ SEQUENCE 360 AA; 38445 MW; C42D1911CB57244A CRC64;

Query Match 74.0%; Score 37; DB 2; Length 360;
Best Local Similarity 88.9%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FAAGAAAG 10
DB 244 FAAGAAAG 252
|||||

RESULT 8

Q9LD54 PRELIMINARY; PRT; 381 AA.
AC Q9LD54;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE ESTS AU068633.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0453A06."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0453A06."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001383; BAA92520.1; -;
DR EMBL; AF001080; BAA90348.1; -;
SQ SEQUENCE 381 AA; 40761 MW; F3A0E3CEBD950778 CRC64;

Query Match 74.0%; Score 37; DB 10; Length 381;
Best Local Similarity 72.7%; Pred. No. 1.9e+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 VFAAGAAAGV 11
DB 92 LFAAGAAAGAV 102
:|||||

```

RESULT 9
O53358 ID O53358 PRELIMINARY; PRT; 394 AA.
AC O53358;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE PUTATIVE AMINOHYDROLASE.
GN RV3306C OR MT016.05C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV.
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
DR EMBL; AL021841; CAA17078.1;
DR TUBERCULIST; RV3306C;
DR INTERPRO; IPR002933;
DR PFAM; PF01546; Peptidase_M20; 2.
KW Hydrolase.
SQ SEQUENCE 394 AA; 40740 MW; 184E98E50534559E CRC64;

```

```

Query Match 74.0%; Score 37; DB 2; Length 394;
Best Local Similarity 88.9%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 2 FAAGAAAAG 10
|||||
Db 259 FAAGALAAAG 267

```

```

RESULT 10
O9RKH6 ID O9RKH6 PRELIMINARY; PRT; 417 AA.
AC O9RKH6;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE PUTATIVE LIPOPROTEIN.
GN SCE46.13C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Brown S.P., Harris D.;
RA Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrrell B.G., Rajandream M.A.;
RA Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;

```

```

RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL133252; CAB61864.1; -.
KW Lipoprotein.
SQ SEQUENCE 417 AA; 45153 MW; FC33BCC3C65F5E46 CRC64;

```

```

Query Match 74.0%; Score 37; DB 2; Length 417;
Best Local Similarity 72.7%; Pred. No. 2.1e+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 VFAAGAAAAGV 11
|||||
Db 13 VLAAGAAVAGL 23

```

```

RESULT 11
O83796 ID O83796 PRELIMINARY; PRT; 630 AA.
AC O83796;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE TRANSKETOLASE B (TKTB).
GN TP0824.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Attiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RA "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
DR EMBL; AE001253; AAC65792.1; -.
DR TIGR; TP0824; -.
SQ SEQUENCE 630 AA; 68102 MW; 70F4E3340E682448 CRC64;

```

```

Query Match 74.0%; Score 37; DB 2; Length 630;
Best Local Similarity 80.0%; Pred. No. 3.1e+02;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 2 FAAGAAAAGV 11
|||||
Db 373 FAAGLACAGV 382

```

```

RESULT 12
O33957 ID O33957 PRELIMINARY; PRT; 1611 AA.
AC O33957;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE TYLACTONE SYNTHASE MODULE 6.
GN TVIG.
OS Streptomyces fradiae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1906;
RN [1]
RP SEQUENCE FROM N.A.
RA DeHoff B.S., Sutton K.L., Rostock P.R. Jr.;

```


RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; 078289; AAB66507.1; -
 DR INTERPRO; IPR000255; -
 DR INTERPRO; IPR000794; -
 DR INTERPRO; IPR001227; -
 DR INTERPRO; IPR002106; -
 DR PFAM; PF00109; ketoacyl-synt; 1.
 DR PFAM; PF00550; pp-binding; 1.
 DR PFAM; PF00698; Acyl_transf; 1.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_1.
 DR PROSITE; PS00339; AA_TRNA_LIGASE_II.2; UNKNOWN_1.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
 DR PROSITE; PS50075; ACP_DOMAIN; 1.
 KW Transferase.
 SQ SEQUENCE 1611 AA; 167436 MW; 315CA090825C29CC CRC64;

Query Match 74.0%; Score 37; DB 2; Length 1611;
 Best Local Similarity 80.0%; Pred. No. 7.5e+02;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VFAAGAAAG 10
 :|||||
 Db 1415 LFAAGFAAG 1424

RESULT 13
 Q9S2I3 PRELIMINARY; PRT; 299 AA.
 AC Q9S2I3;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
 DE HYPOTHETICAL 33.4 KDA PROTEIN.
 GN SC7H2.34.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Saunders D.C., Harris D.;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Redenbach M., Kieser H.M., Denapait D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmid and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RT MOL. Microbiol. 21:77-96(1996).
 DR EMBL; AL109732; CAB52076.1; -
 DR INTERPRO; IPR000130; -
 DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 299 AA; 33418 MW; F6A8A5CDFBF86AD2 CRC64;

Query Match 72.0%; Score 36; DB 2; Length 299;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 3 AAGAAAG 10
 :|||||
 Db 243 AAGAAAG 250

Query Match 72.0%; Score 36; DB 2; Length 299;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 AAGAAAG 10
 :|||||
 Db 243 AAGAAAG 250

RESULT 14
 O69949 PRELIMINARY; PRT; 363 AA.
 AC O69949;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE HYPOTHETICAL 35.1 KDA PROTEIN.
 GN SC3F9.09.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Seeger K.J., Harris D.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Redenbach M., Kieser H.M., Denapait D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmid and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RT MOL. Microbiol. 21:77-96(1996).
 DR EMBL; AL023862; CAA19632.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 363 AA; 35089 MW; 7A4415FB94526FBB CRC64;

Query Match 72.0%; Score 36; DB 2; Length 363;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 AGAAAGV 11
 :|||||
 Db 76 AGAAAGV 83

RESULT 15
 Q9PCP0 PRELIMINARY; PRT; 441 AA.
 AC Q9PCP0;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE HYPOTHETICAL PROTEIN XF1738.
 GN XF1738.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 OX NCBI_TaxID=2371;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvares R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carriaro D.M., Carer H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,

RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Paixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de-Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-157(2000).
DR EMBL: AE003997; AAF84547.1; -.
KW Hypothetical protein.
SQ SEQUENCE 441 AA; 47931 MW; BBDD996A58EC4E89 CRC64;

Query Match 72.0%; Score 36; DB 2; Length 441;
Best Local Similarity 63.6%; Pred. No. 3.1e+02;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 VFAAGAAAGV 11
Db 160 LFAAGATAGGL 170
:|||||I|:
|

Search completed: March 6, 2001, 12:53:02
Job time: 302 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2001, 13:11:11 ; Search time 1324.24 Seconds
(without alignments)
3271.030 Million cell updates/sec

Title: US-09-196-161d-9
Perfect score: 316
Sequence: 1 ggaatcgctcaggagaagc.....tgctgcgctgcaggaaattc 316

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 6853842396 residues 2236266
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_ov.*
- 4: gb_ov.*
- 5: gb_ph.*
- 6: gb_ph.*
- 7: gb_pl1.*
- 8: gb_pl2.*
- 9: gb_pr1.*
- 10: gb_pr2.*
- 11: gb_pr3.*
- 12: gb_ro.*
- 13: gb_sy.*
- 14: gb_sy.*
- 15: gb_un.*
- 16: em_fun.*
- 17: em_hum1.*
- 18: em_hum2.*
- 19: em_in.*
- 20: em_in.*
- 21: em_or.*
- 22: em_or.*
- 23: em_ov.*
- 24: em_ov.*
- 25: em_ph.*
- 26: em_ph.*
- 27: em_pl.*
- 28: em_pl.*
- 29: em_ro.*
- 30: em_ro.*
- 31: em_sts.*
- 32: em_sts.*
- 33: em_sy.*
- 34: em_sy.*
- 35: em_un.*
- 36: em_un.*
- 37: em_v1.*
- 38: gb_htg1.*
- 39: gb_htg2.*
- 40: gb_in1.*
- 41: gb_in2.*
- 42: gb_ba1.*
- 43: gb_ba2.*
- 44: em_hum3.*
- 45: em_hum4.*
- 46: gb_pr4.*
- 47: gb_htg3.*
- 48: gb_htg4.*
- 49: gb_htg5.*
- 50: gb_htg6.*
- 51: gb_htg7.*
- 52: em_htg1.*

- 44: em_htg2.*
- 45: em_htg3.*
- 46: em_hum5.*
- 47: gb_pl3.*
- 48: gb_pr5.*
- 49: gb_htg8.*
- 50: gb_htg9.*
- 51: gb_htg10.*
- 52: gb_htg11.*
- 53: gb_htg12.*
- 54: gb_htg13.*
- 55: gb_htg14.*
- 56: gb_in3.*
- 57: gb_htg15.*
- 58: gb_htg16.*
- 59: gb_htg17.*
- 60: em_htg4.*
- 61: em_htg5.*
- 62: em_htg6.*
- 63: em_htg7.*
- 64: em_hum6.*
- 65: gb_htg18.*
- 66: gb_htg19.*
- 67: gb_htg20.*
- 68: gb_htg21.*
- 69: gb_htg22.*
- 70: gb_htg23.*
- 71: gb_v11.*
- 72: gb_v12.*
- 73: gb_ba3.*
- 74: em_htg8.*
- 75: em_htg9.*
- 76: em_htg10.*
- 77: gb_pr6.*
- 78: gb_pr7.*
- 79: gb_sts1.*
- 80: gb_sts2.*
- 81: gb_pat1.*
- 82: gb_pat2.*
- 83: em_htg0.*
- 84: gb_htg24.*
- 85: gb_pr8.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	283	89.6	1249	56	ICYMANT	M92907 Ichthyophth
2	283	89.6	2486	32	AF140273	AF140273 Ichthyoph
c	43.6	13.8	172307	55	AC044842	AC044842 Homo sapi
3	40.2	12.7	624	32	AF139019	AF139019 Cepaea ne
4	40.2	12.7	22143	69	AL365272	AL365272 Homo sapi
5	39.4	12.5	80009	32	AF226688	AF226688 Bombyx mo
6	38.6	12.2	2384	2	CACSPAG	Z37723 C.acetobuty
7	38.6	12.2	209973	59	AC073755	AC073755 Mus muscu
8	38.4	12.2	1968	11	MMU46463	U46463 Mus musculu
c	38.2	12.1	175302	42	AC020587	AC020587 Homo sapi
9	37.6	11.9	471	78	HSU23863	U23863 Human clone
c	37.4	11.8	22527	40	AC014525	AC014525 Drosophil
10	37.4	11.8	43891	51	AC023719	AC023719 Drosophil
11	37.4	11.8	303092	31	AE003446	AE003446 Drosophil
12	37.2	11.8	171350	55	AC058816	AC058816 Homo sapi
13	37.2	11.7	1722	3	AB048335	AB048335 Equus cab
c	37.0	11.7	41507	40	AC014235	AC014235 Drosophil
14	37.0	11.7	182009	65	AC079031	AC079031 Homo sapi
15	37.0	11.7	310364	31	AE003448	AE003448 Drosophil
c	36.4	11.5	1925	56	PFAANTMMS	J03975 P.yocelli ma
16	36.4	11.5	5774	56	PFAAMSAAC	J04668 P.yocelli ma

```

c 22 36.4 11.5 185994 29 AC002042
c 23 36.4 11.5 196929 59 AC073946
c 24 36.2 11.5 9714 73 U67594
c 25 36.2 11.5 116792 48 HS1100H13
c 26 35.8 11.3 5120 37 AF152102
c 27 35.8 11.3 105922 84 DMR221N6
c 28 35.8 11.3 138097 54 AC027787
c 29 35.6 11.3 3941 11 RNIAR2
c 30 35.6 11.3 133889 77 HS769D20
c 31 35.4 11.2 380 11 RATSIMPB
c 32 35.4 11.2 2126 11 MMU08091
c 33 35.4 11.2 9282 32 AF083334
c 34 35.4 11.2 165822 10 AC009784
c 35 35.4 11.2 210342 42 AC018695
c 36 35.2 11.1 594 11 MUSOPA
c 37 35.2 11.1 1442 10 AF087653
c 38 35.2 11.1 1586 11 MMU70653
c 39 35.2 11.1 1598 11 MMU70654
c 40 35.2 11.1 1612 11 MMU70652
c 41 35.2 11.1 6478 11 AF071310
c 42 35.2 11.1 70878 38 AC010253
c 43 35.2 11.1 110000 39 AC012607_1
c 44 35.2 11.1 180220 41 AC017043
c 45 35 11.1 633 78 HSU23862

```

ALIGNMENTS

```

RESULT 1
ICYMANT 1249 bp mRNA INV 18-SEP-1998
LOCUS Ichthyophthirius multifiliis immobilization antigen precursor,
DEFINITION mRNA, partial cds.
ACCESSION M92907
VERSION M92907.1 GI:3628568
SOURCE Ichthyophthirius multifiliis.
ORGANISM Ichthyophthirius multifiliis.
REFERENCE 1 (bases 1 to 1249)
AUTHORS Clark, R.G., McGraw, R.A. and Dickerson, H.W.
TITLE Developmental expression of surface antigen genes in the parasitic
JOURNAL ciliate Ichthyophthirius multifiliis
MEDLINE Proc. Natl. Acad. Sci. U.S.A. 89, 6363-6367 (1992)
REFERENCE 2 (bases 1 to 1249)
AUTHORS Lin, T.L. and Dickerson, H.W.
TITLE Purification and partial characterization of immobilization
JOURNAL antigens from Ichthyophthirius multifiliis
MEDLINE J. Protozool. 39 (4), 457-463 (1992)
REFERENCE 3 (bases 1 to 1249)
AUTHORS Clark, R.
TITLE Direct Submission
JOURNAL Submitted (06-JUN-1992) Microbiology and Immunology, Cornell
REFERENCE 4 (bases 1 to 1249)
AUTHORS University, Ithaca, NY 14853, USA
TITLE Clark, R.
JOURNAL Direct Submission
REFERENCE 5 (bases 1 to 1249)
AUTHORS University, Ithaca, NY 14853, USA
TITLE Clark, R.
JOURNAL Direct Submission
REMARK Submitted (18-SEP-1998) Microbiology and Immunology, Cornell
COMMENT Sequence update by submitter
FEATURES
    Location/Qualifiers
        1..1249
            /organism="Ichthyophthirius multifiliis"
            /strain="Georgia"
            /db_xref="taxon:5932"
            /dev_stage="trophont"
            <1..1190
            /note="48 kba"

```

CDS

```

/citation=[2]
/citation=[1]
/codon_start=3
/transl_table=6
/evidence=experimental
/product="immobilization antigen precursor"
/protein_id="AAC36158.1"
/db_xref="GI:3628569"
/translaton="AVPCPDGTQAGLTDVGAADLTGVCNCRPNFYNGAAQGEAN
GNOPFAANNAARGICVPCOINRYSVTNAGDLATLATOCSTGCTGTALDDGYDVPD
RSNAQCVCCKPNFYNGGSPQGEAPGVQVFAAGAAAGAAVTSQCVPCLNKNDSPA
TAGAANLATQCSNOCPTGVLDDGVTLVFNSTATLCVKCRPNFYNGSPQGEAPGV
OVFAAGAAAGAAVTSQCVPCLNKNDSPATAGANLATQCSNOCPTGTAQDQGV
LVFNSTQCSNOCIANVFNENLEAGSKCLKCPVSKTTPAHAPGNATATATCCLTTC
PAGTLLDDGTSTNFVASATECTKCSAGFFASKTTGTAGTDTCTECTKLTSGATAVN
IHQ"
sig_peptide 61..5
mat_peptide 61..1187
/evidence=experimental
/product="immobilization antigen"
repeat_region 177..893
/citation=[1]
/rpt_type=tandem
BASE COUNT 348 a 243 c 257 g 401 t
ORIGIN

Query Match 89.6%; Score 283; DB 56; Length 1249;
Best Local Similarity 93.7%; Pred. No. 5.3e-69;
Matches 295; Conservative 0; Mismatches 20; Indels 0; Caps 0;

QY 1 gqatccgcaggaggaagctaataatcgatcagccttccagcaataatgctgctaga 60
Db 111 GGTGCTGCTTAAAGGAGAGCTAATGTAATACCTTTCGACGCAATATGCTGCTAGA 170

QY 61 ggtatatgttaccatgccaaataaacagagtaggctctggtaccacagcaggtagctta 120
Db 171 GGTATATGTGTACCATGCCAAATAAACAGAGTAGGCTCTGTACCAATGCGAGGTGACTTA 230

QY 121 gctactttagccacacaaatcagctactcagtgctcctactggcactgacitgatgaga 180
Db 231 GCACATGTTTATGATAGATAGATGCTTAAATGCTTACTGCTGCTGCTGCTGCTGCTGCT 290

QY 181 gtacacagatgtttttagatagatcagccgacagtggttaataatgcaacacacacacacacac 240
Db 291 GTGACAGATGTTTATGATAGATAGATGCTTAAATGCTTAAATGCAACACACACACACACAC 350

QY 241 tataatggtggttctcctcaggtggaagctcctggtcctggtcctggttctggtggtgct 300
Db 351 TATAATGTTGTTCTCCTTAAAGGTGAAGCTCCTGGCGCTTAAAGTTTGTGCTGCTGCTGCT 410

QY 301 gccgctgacaggaatt 315
Db 411 GCCGCTGACAGGTGT 425

RESULT 2
AF140273
LOCUS Ichthyophthirius multifiliis INV 19-MAY-1999
DEFINITION (IAG48) gene, complete cds.
ACCESSION AF140273
VERSION AF140273.1 GI:4868370
KEYWORDS Ichthyophthirius multifiliis.
SOURCE Ichthyophthirius multifiliis.
ORGANISM Ichthyophthirius multifiliis.
REFERENCE 1 (bases 1 to 2486)
AUTHORS Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
Hymenostomatida; Ophryoglenina; Ichthyophthirius.
Clark, T.G., Lin, T.L., Jackwood, D.A., Sherrill, J., Lin, Y. and
Dickerson, H.W.
TITLE The gene for an abundant parasite coat protein predicts tandemly

```


* arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 19: contig of 19 bp in length
20 119: gap of 100 bp
120 1921: contig of 1802 bp in length
1922 2021: gap of 100 bp
2022 3689: contig of 1668 bp in length
3690 3789: gap of 100 bp
3790 6531: contig of 2742 bp in length
6532 6631: gap of 100 bp
6632 8134: contig of 1503 bp in length
8135 8234: gap of 100 bp
8235 11919: contig of 3685 bp in length
11920 12019: gap of 100 bp
12020 15244: contig of 3225 bp in length
15245 15344: gap of 100 bp
15345 17344: contig of 2000 bp in length
17345 17444: gap of 100 bp
17445 21830: contig of 4386 bp in length
21831 21930: gap of 100 bp
21931 25081: contig of 3151 bp in length
25082 25181: gap of 100 bp
25182 29736: contig of 4555 bp in length
29737 29836: gap of 100 bp
29837 36853: contig of 7017 bp in length
36854 36953: gap of 100 bp
36954 41270: contig of 4317 bp in length
41271 41370: gap of 100 bp
41371 46765: contig of 5395 bp in length
46766 46865: gap of 100 bp
46866 57601: contig of 10736 bp in length
57602 57701: gap of 100 bp
57702 67879: contig of 10178 bp in length
67880 67979: gap of 100 bp
67980 78235: contig of 10256 bp in length
78236 78335: gap of 100 bp
78336 89176: contig of 10841 bp in length
89177 89276: gap of 100 bp
89277 106081: contig of 16805 bp in length
106082 106181: gap of 100 bp
106182 121387: contig of 15206 bp in length
121388 121487: gap of 100 bp
121488 137207: contig of 15720 bp in length
137208 137307: gap of 100 bp
137308 153745: contig of 16438 bp in length
153746 153845: gap of 100 bp
153846 172307: contig of 18462 bp in length.
  
```

FEATURES

```

source
  1. .172307
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /map="2"
    /chromosome="2"
    /clone="RP11-178E20"
    /clone_lib="RPC1-11 Human Male BAC"
  1. .19
    /note="assembly_fragment"
    clone_end:77
    vector_side:left
  120. .1921
    /note="assembly_fragment"
  2022. .3689
    /note="assembly_fragment"
  3790. .6531
    /note="assembly_fragment"
  6632. .8134
    /note="assembly_fragment"
  8235. .11919
    /note="assembly_fragment"
  12020. .15244
  
```

```

misc_feature
  15345. .17344
    /note="assembly_fragment"
  clone_end:SP6
  vector_side:right
  17445. .21830
    /note="assembly_fragment"
  21931. .25081
    /note="assembly_fragment"
  25182. .29736
    /note="assembly_fragment"
  29837. .36853
    /note="assembly_fragment"
  36954. .41270
    /note="assembly_fragment"
  41371. .46765
    /note="assembly_fragment"
  46866. .57601
    /note="assembly_fragment"
  57702. .67879
    /note="assembly_fragment"
  67980. .78235
    /note="assembly_fragment"
  78336. .89176
    /note="assembly_fragment"
  89277. .106081
    /note="assembly_fragment"
  106182. .121387
    /note="assembly_fragment"
  121488. .137207
    /note="assembly_fragment"
  137308. .153745
    /note="assembly_fragment"
  153846. .172307
    /note="assembly_fragment"
  44249 a 40378 c 41592 g 43880 t 2208 others
  ORIGIN

Query Match      13.8%; Score 43.6; DB 55; Length 172307;
Best Local Similarity 46.9%; Pred. No. 0.077;
Matches 136; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 19 gctaagtgaatcagccttgcagcaataatgctgctagaggtatgtgtaccatgc 78
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38858 GCTGATGCTACTCTCTCTGCCACAGCTACTGCAACTACTGATGCTGCTCTGCC 38799

QY 79 caaataacagagtagctctgtgtacccaatgcaggtgacttagctacttagccacaa 138
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38798 ACAGCTACTGCAACTGCTGATGCTGCTACTCTCTGCCACAGCTACTGCAACTGCT 38739

QY 139 tgcagtactcagtgctctactggcactgacattgatgagtgacagatgttttgat 198
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38738 GCTGCTACTCTCTCTGCCACAGCTACTGCAACAGCTGCTGCTGCTACTCTGCCACAGCT 38679

QY 199 agatcagccgacagtggtgttaaatgcaaacctacttactactaaatgggtgtctcct 258
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38678 ACTGCAACTGCTGCTACTCTCTCTGCCACAGCTACTTCAACTGCTGATGCTGCTACTCTCT 38619

QY 259 cagggtgaagctcctgccttcagggttttgcgtgctggtgctgcgcgtgc 308
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38618 CCTGCCACAGCTACTGCAACTGCTGATGCTGCTACTCTCTCTGCCACAGC 38569

RESULT 4
AF139019
LOCUS      AF139019      624 bp      DNA      INV      05-AUG-1999
DEFINITION Cepaea nemoralis microsatellite Cnel sequence.
ACCESSION  AF139019
VERSION     AF139019.1  GI:5702210
KEYWORDS   banded wood snail.
SOURCE     Cepaea nemoralis
  
```



```

repeat_region 277.306 /rpl_type=tandem
repeat_region /rpl_family="opa (CAZ repeat)"
3'UTR 310.624 /rpl_type=tandem
polyA_signal /rpl_family="opa (CAZ repeat)"
polyA_site 697.1968
1928.1933
BASE COUNT 615 a 475 c 428 g 450 t
ORIGIN

Query Match 12.2% Score 38.4; DB 11; Length 1968;
Best Local Similarity 51.8%; Pred. No. 2.4;
Matches 87; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 143 gtaactagttcctactgacactgacattgatgagtgacagatctttttagatagat 202
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 552 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 493

QY 203 cagccgacagtggttaaatgcaaacactacttactataatggtgtctctcagg 262
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 492 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 433

QY 263 gtgaagctctgacctcaggttttgtctgtgtgtgtgtgtgtgtgtgtgtgtgt 310
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 432 CTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 385

RESULT 10
AC020587/c AC020587 175302 bp DNA HTG 17-AUG-2000
LOCUS Homo sapiens chromosome 1 clone RP11-99F3, WORKING DRAFT SEQUENCE,
DEFINITION 21 unordered pieces.
ACCESSION AC020587.4 GI:9838066
VERSION AC020587.4
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 175302)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 175302)
REFERENCE Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (05-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Aug 17, 2000 this sequence version replaced gi:7637330.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information:
Center project name: H_NH0099F03
----- Summary Statistics -----
Sequencing vector: M13; 84%
Chemistry: Dye-terminator Big Dye; 16%
Assembly program: Phrap; version 0.990319
Consensus quality: 162600 bases at least Q40
Consensus quality: 166672 bases at least Q30
Consensus quality: 168705 bases at least Q20
Insert size: 183000; agarose-fp
Insert size: 174443; sum-of-contigs
Quality coverage: 4.09 in Q20 bases; agarose-fp
Quality coverage: 4.37 in Q20 bases; sum-of-contigs

```

```

-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1486: contig of 1486 bp in length
* 1487 1586: gap of unknown length
* 1587 3436: contig of 1850 bp in length
* 3437 3536: gap of unknown length
* 3537 6233: contig of 2697 bp in length
* 6234 6334: gap of unknown length
* 6334 8454: contig of 2121 bp in length
* 8455 8555: gap of unknown length
* 8555 10600: contig of 2046 bp in length
* 10601 10701: gap of unknown length
* 10701 13742: contig of 3042 bp in length
* 13743 13843: gap of unknown length
* 13843 18085: contig of 4243 bp in length
* 18086 18186: gap of unknown length
* 18186 22540: contig of 4355 bp in length
* 22541 22641: gap of unknown length
* 22641 27125: contig of 4485 bp in length
* 27126 27226: gap of unknown length
* 27226 31677: contig of 4451 bp in length
* 31677 31777: gap of unknown length
* 31777 37650: contig of 5874 bp in length
* 37651 37750: gap of unknown length
* 37751 46532: contig of 8782 bp in length
* 46533 46632: gap of unknown length
* 46633 52522: contig of 5889 bp in length
* 52522 52622: gap of unknown length
* 52622 62167: contig of 9546 bp in length
* 62168 62267: gap of unknown length
* 62268 71200: contig of 8933 bp in length
* 71201 71300: gap of unknown length
* 71301 80071: contig of 8771 bp in length
* 80072 80171: gap of unknown length
* 80172 88991: contig of 8820 bp in length
* 88992 89091: gap of unknown length
* 89092 100957: contig of 11866 bp in length
* 100958 101057: gap of unknown length
* 101058 113645: contig of 12588 bp in length
* 113646 113745: gap of unknown length
* 113746 136276: contig of 22531 bp in length
* 136277 136376: gap of unknown length
* 136377 175302: contig of 38926 bp in length.
*
* Location/Qualifiers
* 1..175302
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="1"
* /clone="RP11-99F3"
* 1..1486
* /note="assembly_name:Contig12"
* 1587..3436
* /note="assembly_name:Contig14"
* 3537..6233
* /note="assembly_name:Contig15"
* 6334..8454
* /note="assembly_name:Contig16"
* 8555..10600
* /note="assembly_name:Contig18"
* 10701..13742
* /note="assembly_name:Contig19"
* 13843..18085
* /note="assembly_name:Contig20"
* 18186..22540
* /note="assembly_name:Contig21"
* 22641..27125

```


BASE COUNT	ORIGIN	FEATURES
11694 a	9728 c	9711 g
11943 t	815	others
/organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="RP98-4L24" Location/Qualifiers 1..43891 * 35162: gap of unknown length * 35163 36354: contig of 1192 bp in length * 36355 36374: gap of unknown length * 36375 37957: contig of 1383 bp in length * 37958 37977: gap of unknown length * 37978 39554: contig of 1577 bp in length * 39555 39574: gap of unknown length * 39575 40925: contig of 1351 bp in length * 40926 40945: gap of unknown length * 40946 43891: contig of 2946 bp in length.		

[illegible]

QY 215 gfgttaatgcaactacttltactataatgylgtlctcctcaggytgaagctcctg 274

QY 275 gccctcaggtlcttgcctgctgctgcgcctgcaggaa 313
| ||| ||||| || ||||| ||||| |||
Db 36829 CAACGCAAGTTGCTGCTGATGCTGCTGCTGCGCA 36867

RESULT	14
AE003446	
LOCUS	
DEFINITION	AE003446 303092 bp DNA INV 04-OCT-2000
ACCESSION	Drosophila melanogaster genomic scaffold 142000013386054 section 30 of 35, complete sequence.
VERSION	AE003446.2 GI:10728536

ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 303092)
REFERENCE
Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,

REFERENCE
AUTHORS

1 (bases 1 to 303092)

Adams, M.D., Celinker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,
Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.,
George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,
Sutton, G.G., Wortman, J.R., Vandeil, M.D., Zhang, Q., Chen, L.X.,
Brandon, R.C., Rogers, Y.H., Blazer, R.G., Champe, M., Pfeiffer, B.D.,
Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor
Miklos, G.L., Abril, J.F., Agbayani, A., An, H.J.,
Andrews, Plank Koch, C., Baldwin, D., Ballew, R.M., Basu, A.,
Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y.,
Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D.,
Botchan, M.R., Bouck, J.P., Brokstein, P., Brotler, P., Butts, K.C.,
Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I.,
Cherry, J.M., Cawley, S., Dalkile, C., Davenport, L.B., Davies, P.,
de Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M.,
Dodson, K., Dou, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C.,
Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferreira, S.,
Fleischmann, W., Foster, C., Gabrielian, A.E., Garg, N.J.,
Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z.,
Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J.,
Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J.,
Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z.,
Kernison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C.,
Kraivetz, S., Kulp, D., Lai, Z., Lasco, P., Lei, Y., Levitsky, A.,

TITLE The genome sequence of *Drosophila melanogaster*
 JOURNAL Science 287 (5461), 2185-2195 (2000)
 MEDLINE 20196006
 REFERENCE 2 (bases 1 to 303092)
 AUTHORS Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
 TITLE Direct Submission
 JOURNAL Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 COMMENT On Oct 9, 2000 this sequence version replaced gi:7290992.
 FEATURES
 source Location/Qualifiers
 1..303092

mRNA

genes

CDS

miRNA

gene

CDS

Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T. C., McLeod, M. P., McPherson, D., Meriklo, G., Milshina, N. V., Mobarry, C., Morris, J. D., Mosnelli, A., Mount, S. M., Moy, M., Murphy, B., Murphy, L., Muzny, D. M., Nelson, D. L., Nelson, S. M., Nix, M., Nixson, K. A., Nixson, K., Nusskern, D. R., Pacle, J. M., Palazzolo, M., Pittman, G. S., Pan, S., Pollard, J. D., Puri, V., Reese, M. G., Reinert, K., Remington, K., Saunders, R. D., Scheeler, F., Shen, H., Shie, B. C., Siden-Kiamos, I., Simpson, M., Skupski, M. P., Smith, T., Spier, E., Spradling, A. C., Stapleton, M., Strong, R., Sun, E., Sytkas, R., Tector, C., Turner, R., Venter, E., Wang, A. H., Wang, X., Wang, Z. Y., Wasserman, D. A., Weinstein, G. M., Weissensbach, J., Williams, S. M., Woodage, T., Worley, K. C., Wu, D., Yang, S., Yao, Q. A., Ye, J., Yeh, R. F., Zaverii, J. S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X. H., Zhong, F. N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H. O., Gibbs, R. A., Myers, E. W., Rubin, G. M., and Venter, J. C.

The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)

20196006

2 (bases 1 to 303092)

Adams, M. D., Celisner, S. E., Gibbs, R. A., Rubin, G. M. and Venter, C. J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
On Oct 9, 2000 this sequence version replaced gi:7290992.

Location/Qualifiers

1. 303092

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/chromosome="X"

join(<6691..6699,8631..>8972)

/gene="CG15364"

/product="CT35383"

/db_xref="FLYBASE:FBan0015364"

/db_xref="FLYBASE:FBgn0030075"

/evidence=not_experimental

<6691..>8972

/gene="CG15364"

/map="8C7-8C9"

/db_xref="FLYBASE:FBan0015364"

/db_xref="FLYBASE:FBgn0030075"

/evidence=not_experimental

join(<6691..6699,8631..8972)

/gene="CG15364"

/note="CG15364 gene product"

/codon_start=1

/db_xref="FLYBASE:FBan0015364"

/db_xref="FLYBASE:FBgn0030075"

/evidence=not_experimental

/protein_id="AA046431.1"

/db_xref="GI:7200993"

/translation="MYSLSRSARYAKAKSKSAINLNHLKROOKLEEGHPPEKRODK
KILQVSEQDPIYAVDEYPIHLEPETGEVDALMQNAGVSGPKNPQTQOTSSI
CIRHCHPSKIPN"

join(<17819..19495,19597..19662,19745..19970,20028..20777,
20838..20948,21010..21879)

/gene="CG4040"

/product="CT13364"

/db_xref="FLYBASE:FBan0004040"

/db_xref="FLYBASE:FBgn0030076"

<17819..>21879

/gene="CG4040"

/map="8C8-8C9"

/db_xref="FLYBASE:FBan0004040"

/db_xref="FLYBASE:FBgn0030076"

join(<17819..19495,19597..19662,19745..19970,20028..20777,
20838..20948,21010..21879)

/gene="CG4040"

/note="CG4040 gene product"

/codon_start=1

/db_xref="FLYBASE:FBan0004040"

/db_xref="FLYBASE:FBgn0030076"

/protein_id="AA046432.1"

/db_xref="GI:7290994"

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2001, 18:13:31 ; Search time 88.23 Seconds
(without alignments)
1345.454 Million cell updates/sec

Title: US-09-196-161d-9

Perfect score: 316
Sequence: 1 ggaaccgcctcagggaggaagc.....tgctgcctcgcaggaatc 316

Scoring table:
IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_36:*

1: /cgnl_8/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /cgnl_8/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /cgnl_8/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /cgnl_8/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /cgnl_8/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /cgnl_8/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /cgnl_8/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /cgnl_8/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /cgnl_8/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /cgnl_8/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /cgnl_8/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /cgnl_8/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /cgnl_8/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /cgnl_8/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /cgnl_8/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /cgnl_8/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /cgnl_8/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /cgnl_8/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /cgnl_8/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /cgnl_8/gcgdata/geneseq/geneseqn/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	37	11.7	543	13	Q23092
C 2	36.2	11.5	1664976	19	Antigen tc-7a gene
C 3	35.2	11.1	6558	21	Methanococcus jann
C 4	34.8	11.0	6794	21	CDNA sequence of a
5	34.2	10.8	515	20	Human gonadotropin
6	34	10.8	32207	20	Human gonadotropin
7	34	10.8	137507	19	SHV LUR DNA (nucl
C 8	33.8	10.7	580073	18	Myocoplasmia genital
C 9	33.6	10.6	234	16	Myocoplasmia genital
10	33.6	10.6	1185	21	Spinoerebellar at
C 11	33.4	10.6	397	20	Escherichia coli f
12	32.8	10.4	2322	17	Spinoerebellar at
					Malic enzyme codin

C 13	32.6	10.3	1987	20	240460
C 14	32.6	10.3	2769	18	T59196
C 15	32.6	10.3	3041	20	240461
C 16	32.6	10.3	5970	20	X26546
C 17	32.6	10.3	6789	19	V29372
C 18	32.6	10.3	7791	16	O84660
C 19	32.6	10.3	7808	16	O84659
C 20	32.4	10.3	10266	17	T33007
C 21	32.4	10.3	14704	13	O20685
C 22	31.4	9.9	477	21	244307
C 23	31.4	9.9	659	19	V59918
C 24	31.4	9.9	1413	20	X07315
C 25	31.2	9.9	195	16	O84831
C 26	31	9.8	1064	16	O87093
C 27	31	9.8	1064	18	T63677
C 28	31	9.8	1064	20	X03962
C 29	31	9.8	1270	16	O87101
C 30	31	9.8	1270	18	T63685
C 31	31	9.8	1270	20	X03970
C 32	31	9.8	580073	18	T58840
C 33	30.8	9.7	1206	21	258569
C 34	30.6	9.7	1727	16	O91871
C 35	30.6	9.7	10240	19	V39007
C 36	30.4	9.6	1413	20	X07307
C 37	30.4	9.6	1413	20	X07308
C 38	30.4	9.6	1413	20	X07310
C 39	30.2	9.6	203	19	V30271
C 40	30.2	9.6	203	19	V17226
C 41	30.2	9.6	1424	13	Q29273
C 42	30.2	9.6	3569	18	N91772
C 43	30.2	9.5	153	18	T78907
C 44	30	9.5	165	19	V30274
C 45	30	9.5	165	19	V30274

ALIGNMENTS

RESULT 1

Q23092/c

ID Q23092 standard; DNA, 543 BP.

XX Q23092:

XX 17-AUG-1992 (first entry)

DE Antigen tc-7a gene.

XX

XX Oocysts; Mab 12-07; sporozoite; ss.

OS Elmeria tenella.

XX

PN WO9204460-A.

XX

XX 19-MAR-1992.

PD

XX

PF 05-SEP-1991; 91MO-US06430.

XX

PR 12-SEP-1990; 90US-0581693.

XX

PA (GENE-) GENEX CORP.

XX

PI Jacobson JW, Strausberg RL, Wilson SD, Pope SH, Strausberg SL;

PI Ruff MD, Augustine PC, Danforth HD;

XX

DR WPI, 1992-114365/14.

DR P-PSDB; R22392.

XX

PT Vaccine against avian coccidiosis - comprising recombinant

PT Elmeria antigen ac-1b or ac-6b gene, or microorganisms expressing

PT them

XX

Claim 5; Page 39 + Fig 5; 56pp; English.

[illegible]

PD		04-NOV-1999.	
PF	xx		
PR	xx	29-APR-1999;	99MO-USO0365.
XX	xx	29-APR-1998;	98US-0083465.
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.		
PA	(IOWA) UNIV IOWA RES FOUND.		
PI	Philbert RA, Gims EI;		
DR	WPI; 2000-126357/11.		
PT	Identification of polymorphisms in the PCTG4 region of Xq13 for		
PS	diagnosing mental retardation or autism .		
XX			
XX	Example 7: Page 68-71; 100pp: English.		
XX			
CC	The present sequence represents the cDNA sequence of the human OPA		
CC	gene, which is derived from the human PCTG4 region of chromosome Xq13		
CC	polymorphisms in this region are associated with mental retardation,		
CC	autism, depression, bipolar affective disorder or hypothyroidism.		
CC	One 12 bp insertion polymorphism occurs within the coding region of		
CC	the human OPA gene, and introduces a 4 amino acid insertion in a		
CC	putative OPA domain. This domain has been shown to be involved in tissue		
CC	specific expression. Another polymorphism consists of a pentanucleotide		
CC	repeat approximately 7 kb upstream of the 12 bp polymorphism. Another		
CC	polymorphism consists of a dinucleotide repeat approximately 4.5 kb		
CC	downstream of the 12 bp polymorphism. The specification describes a		
CC	method for screening for polymorphisms in a PCTG4 nucleic acid sequence		
CC	obtained from a subject. The PCTG4 related sequences within the q13		
CC	region of the x chromosome have polymorphisms associated with		
CC	neuropsychiatric disorders. The methods can be used to screen for the		
CC	presence of a heritably linked form of mental retardation, autism,		
CC	depression, bipolar affective disorder or hypothyroidism.		
SQ	Sequence 6794 BP; 1624 A; 1938 C; 1754 G; 1478 T; 0 other:		
Query Match	11.0%; Score 34.8; DB 21; Length 6794;		
Best Local Similarity	51.3%; Pred. No. 0.46;		
Matches 81; Conservative	0; Mismatches 77; Indels	0; Gaps	
Db	61323 TGCTGTGCCTGCTCTGCTCTGCTTCGTGGTCTCCTCGACGAGATCTGCTGCTGC	6266	
OY	213 gtgtgtaaatgcacaaacttaacttataatagtgttcctcccaagtggaagctcc	272	
Db	6263 TGCTGTGCCGAGATGATGTAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGC	6204	
OY	273 tggcctcaggtttttgtctgctgtagtgcgcgtcgag	310	
Db	6203 TGCTGTGCTGTTGCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	6166	
RESULT	5		
ID	240457		
XX	240457 standard; cDNA; 515 BP.		
AC	240457;		
DT	15-FEB-2000 (first entry)		
DE	Human gonadotropin receptor partial coding sequence #1.		
KM	Human; gonadotropin receptor; screening; ligand; biomedical research;		
KW	biochemical research; drug; hormone; reproductive tissue; infertility;		
OS	contraction; ss.		
XX	Homo sapiens.		
XX	EP950711-A2		

XX 20-OCT-1999.
 XX 02-FEB-1999; 99EP-0200303.
 XX 06-FEB-1998; 98EP-0200357.
 PR 27-JUL-1998; 98EP-0202519.
 PR 24-SEP-1998; 98EP-0203213.
 XX (ALKU) AKZO NOBEL NV.
 XX PI Van Der Spek PJ, Heikoop JC;
 XX WPI; 1999-563673/48.
 DR P-PSDB; Y53571.
 XX
 PT New 7 transmembrane gonadotropin receptors, useful for screening for
 PT hormone analogs and drugs -
 PS Claim 2; Page 9; 38pp; English.
 XX
 CC Sequences 240457-240464 represent fragments of 3 novel human gonadotropin
 CC receptor genes. The novel gonadotropin receptors can be used to screen
 CC for ligands of the receptors. This screen may be used in biomedical and
 CC biochemical research to develop new drugs targeted to the gonadotropin
 CC receptors. For example, hormone analogs which activate or inhibit the
 CC function of the gonadotropin receptors or classical gonadotropin
 CC receptors may be detected. The altered expression or dysfunction of
 CC the gonadotropin receptors causes conditions in reproductive tissues.
 CC The ligands can be used for the treatment of infertility or for
 CC contraception.
 XX
 SQ Sequence 515 BP; 65 A; 181 C; 148 G; 121 T; 0 other;

Query Match 10.8%; Score 34.2; DB 20; Length 515;
 Best Local Similarity 62.1%; Pred. No. 0.23; Mismatches 33; Indels 0; Gaps 0;
 Matches 54; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 224 gcaaacctacttactataatggtctctccacaggtgaagctcctgacctcaag 283
 II IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
 DB 152 gacctcctcagcttgctccatgctggtgacctcttccctgtcaagccgagccgtcaag 211
 QY 284 ttttctgctgtgtgctgcgcgtgcag 310
 I
 DB 212 tctgtcctgctgtgtgtgtcgtccctg 238

RESULT 6
 V73805
 ID V73805 standard; DNA; 32207 BP.
 XX
 AC V73805;
 XX
 DT 25-FEB-1999 (first entry)
 XX
 DE KSHV LUR DNA (nucleotides 105,301-137,507).
 XX
 KW Kaposi's sarcoma; acquired immune deficiency syndrome; AIDS; DHFR; Bcl-2;
 KW dihydrofolate reductase; LUR; long unique region; vaccine; prophylaxis;
 KW diagnosis; treatment; HHV8; capsid protein IV; tegument protein IV;
 KW glycoprotein; Kaposin; cyclin D; immediate early protein; IEP; OX-2;
 KW v-adh; G-protein coupled receptor; FGARAT; ds.
 XX
 OS Kaposi's sarcoma-associated herpesvirus.
 XX
 PN US5849564-A.
 XX
 PD 15-DEC-1998.
 XX
 PF 29-NOV-1996; 96US-0770379.
 XX
 PR 29-NOV-1996; 96US-0770379.

XX (UYCO) UNIV COLUMBIA NEW YORK.
 PA Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;
 PI WPI; 1999-069741/06.
 DR
 XX
 XX Kaposi's sarcoma-associated herpes virus nucleic acid - encodes
 PT dihydrofolate reductase and is useful for treatment, prophylaxis
 PT or diagnosis of Kaposi's sarcoma
 XX
 PS Disclosure; Column 155-182; 109pp; English.
 XX
 CC This sequence is a fragment of the Kaposi's sarcoma-associated
 CC herpesvirus (KSHV) LUR (long unique region). This fragment contains
 CC coding regions for ORF65 which encodes capsid protein IV, ORF66, ORF67
 CC which encodes tegument protein IV, ORF68 which encodes a glycoprotein,
 CC ORF69, K12 which encodes Kaposin, K13, ORF72 which encodes a cyclin D,
 CC ORF73 which encodes immediate early protein (IEP), K14 which encodes
 CC OX-2 (v-adh), ORF74 which encodes G-protein coupled receptor, ORF75
 CC which encodes tegument protein/FGARAT, K15. KSHV is a new human
 CC Herpesvirus (HHV8) believed to cause Kaposi's sarcoma (KS) which is the
 CC most common form of neoplasm occurring in persons with acquired immune
 CC deficiency syndrome (AIDS). The DHFR protein is useful for vaccination,
 CC prophylaxis, diagnosis and treatment of a subject with Kaposi's sarcoma
 CC and for detecting expression of a DNA virus associated with Kaposi's
 CC sarcoma in a cell.
 XX
 SQ Sequence 32207 BP; 7229 A; 9156 C; 8713 G; 7109 T; 0 other;

Query Match 10.8%; Score 34; DB 20; Length 32207;
 Best Local Similarity 44.4%; Pred. No. 1.6; Mismatches 136; Conservative 0; Mismatches 170; Indels 0; Gaps 0;
 Matches 136; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 5 ccgtcagggaagaagcctaagttaatacgccttcgcagcaataatgctgtagagta 64
 I
 DB 19891 ctgctcctgctcaccctcgtcgtctcctcctgctgctgctcaccctcgtcgtcctac 19950
 QY 65 tatgtaccacgacgaataaacaagtagtagctgtgtaccatgacaggtgacttagta 124
 II
 DB 19951 ctgctgctgctcaccctcgtcgtcgtcctacccctgctgctcaccctcgtcgtcctac 20010
 QY 125 cttagcacaacaatgacagtaactgctcctcactgacagctgacatgtagatga 184
 II
 DB 20011 ctgctgctgctcaccctcgtcgtcgtcctacccctgctgctcaccctcgtcgtcctac 20070
 QY 185 cagatgtttttgataagatcagcgacagtggttaaatgacaaccttaactata 244
 I
 DB 20071 ctgctgctgctcaccctcgtcgtcgtcctacccctgctgctcaccctcgtcgtcctac 20130
 QY 245 atggtgttctcctcaggaagctcctgacctcgaagtttggctggtgctgcg 304
 II
 DB 20131 ctgctgctgctcaccctcgtcgtcgtcctacccctgctgctcaccctcgtcgtcctac 20190
 QY 305 ctgcag 310
 I I I I I
 DB 20191 ctgctg 20196

RESULT 7
 V19941
 ID V19941 standard; DNA; 137507 BP.
 XX
 AC V19941;
 XX
 DT 03-AUG-1998 (first entry)
 XX
 DE KSHV long unique coding region and terminal repeat.
 XX
 KW KSHV; HHV8; human herpes virus 8; macrophage inflammatory protein II;
 KW interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis;
 KW complement-binding protein; glycoprotein; capsid protein IV; infection;

KW	M. genitalium; DNAA: DNase gyrase; origin of replication;
KX	megabase shotgun sequencing method; open reading frame; ORF; SS.
OS	Mycoplasma genitalium.
XK	
FH	Location/Qualifiers
CDS	8552..9184 /*tag= a /label= MG006 /note= "Previously identified as MORF-20076, the encoded protein shows 27.59 percentage identity to thymidylate kinase (CDC8) from Saccharomyces cerevisiae"
FT	11252..12040 /*tag= b /label= MG009 /note= "Previously identified as MORF-20078, the encoded protein shows 35.43 percentage identity to the Bacillus subtilis hypothetical protein covered in accession number GB:D26185_102"
CDS	12069..12725 /*tag= c /label= MG010 /note= "Previously identified as MORF-20079, the encoded protein shows 25.73 percentage identity to DNA primase (dnae) from Clostridium acetobutylicum"
CDS	complement (13570..14247) /*tag= d /label= MG012 /note= "Previously identified as MORF-20080, the encoded protein shows 31.50 percentage identity to the ribosomal protein S6 modification protein (rimk) from Escherichia coli"
CDS	complement (14396..15217) /*tag= e /label= MG013 /note= "Previously identified as MORF-19823, MORF-20080 and MORF-20081, the encoded protein shows 33.04 percentage identity to 5,10-methylene-tetra-hydrofolate dehydrogenase (fold) from E. coli"
CDS	17474..19243 /*tag= f /label= MG015 /note= "Previously identified as MORF-20084, the encoded protein shows 32.23 percentage identity to transport ATP-binding protein (msdA) from E. coli"
CDS	26478..27344 /*tag= g /label= MG023 /note= "Previously identified as MORF-20092, the encoded protein shows 45.96 percentage identity to fructose-bisphosphate aldolase (tsr) from B. subtilis"
CDS	27345..28448 /*tag= h /label= MG024 /note= "Previously identified as MORF-19926 and MORF-20093, the encoded protein shows 46.84 percentage identity to GTP-binding protein from E. coli"
CDS	36987..38978 /*tag= i /label= MG032 /note= "Previously identified as MORF-20099, the encoded protein shows 26.82 percentage identity to ATP-dependent nuclease (adda) from B. subtilis"
CDS	39242..39904 /*tag= j
FT	/label= MG033 /note= "Previously identified as MORF-20100, the encoded protein shows 35.90 percentage identity to glycerol uptake facilitator (glpf) from B. subtilis"
CDS	complement (39873..40514)
FT	/*tag= k /label= MG034 /note= "Previously identified as MORF-20101, the encoded protein shows 48.13 percentage identity to thymidylate kinase (tdk) from B. subtilis"
CDS	40543..41787 /*tag= l /label= MG035 /note= "Previously identified as MORF-20102, the encoded protein shows 30.71 percentage identity to histidinyl-tRNA synthetase (hisS) from Mycobacterium leprae"
CDS	complement (44751..46277)
FT	/*tag= m /label= MG038 /note= "Previously identified as MORF-20105, the encoded protein shows 46.83 percentage identity to glyceral kinase (gipk) from E. coli"
CDS	complement (46268..47422)
FT	/*tag= n /label= MG039 /note= "Previously identified as MORF-19831 and MORF-20106, the encoded protein shows 43.20 percentage identity to glycerol-3-phosphate dehydrogenase (GUT2) from S. cerevisiae"
CDS	49377..49643 /*tag= o /label= MG041 /note= "The encoded protein shows 48.86 percentage identity to phosphohistidineprotein-leucine phosphoryltransferase (ptsh) from Mycoplasma capricolum"
CDS	50060..51520 /*tag= p /label= MG042 /note= "Previously identified as MORF-19832 and MORF-20108, the encoded protein shows 41.92 percentage identity to spermidine/putrescine transport system permease protein C (potC) from E. coli"
CDS	51525..52382 /*tag= q /label= MG043 /note= "Previously identified as MORF-20110, the encoded protein shows 26.51 percentage identity to spermidine/putrescine transport system permease protein (pcbB) from E. coli"
CDS	52366..53220 /*tag= r /label= MG044 /note= "Previously identified as MORF-20111, the encoded protein shows 29.45 percentage identity to spermidine/putrescine transport system permease protein C (potC) from E. coli"
CDS	54658..55605 /*tag= s /label= MG046 /note= "Previously identified as MORF-20112, the encoded protein shows 36.60 percentage identity to stialoglycoproteinase (gcp) from Pasteurella haemolytica"
CDS	complement (56970..58310)
FT	/*tag= t /label= MG048 /note= "Previously identified as MORF-19834,


```

FT MORF-20114 and MORF-20115, the encoded protein
FT shows 43.02 percentage identity to signal
FT recognition particle protein (fif) from B.
FT subtilis"
FT CDS
FT      58117..59079
FT      /*tag= u
FT      /label= MG049
FT      /note= "Previously identified as MORF-20114 and
FT      MORF-20115, the encoded protein shows 44.78
FT      percentage identity to purine-nucleoside
FT      phosphorylase (deod) from E. coli"
FT      59083..59754
FT      /*tag= v
FT      /label= MG050
FT      /note= "Previously identified as MORF-20117, the
FT      encoded protein shows 83.03 percentage
FT      identity to deoxyribose-phosphate aldolase
FT      (deoc) from Mycoplasma pneumoniae"
FT CDS
FT      complement (64898..65731)
FT      /*tag= w
FT      /label= MG056
FT      /note= "Previously identified as MORF-20122, the
FT      encoded protein shows 30.25 percent
FT      identity to the protein disclosed in
FT      GB:D26185_99 from B. subtilis"
FT CDS
FT      complement (65713..66249)
FT      /*tag= x
FT      /label= MG057
FT      /note= "Previously identified as MORF-20123, the
FT      encoded protein shows 38.90 percentage
FT      identity to the protein disclosed in
FT      GB:D26185_104 from B. subtilis"
FT CDS
FT      81047..82597
FT      /*tag= y
FT      /label= MG067
FT      /note= "Previously identified as MORF-19845, the
FT      encoded protein shows 28.84 percentage
FT      identity to glutamic acid specific protease
FT      (spase) from staphylococcus aureus"
FT CDS
FT      91065..91919
FT      /*tag= z
FT      /label= MG070
FT      /note= "Previously identified as MORF-20136, the
FT      encoded protein shows 34.8 percentage
FT      identity to ribosomal protein S2 (rps2)
FT      from Spirulina plantensis"
FT
FT Query Match          10.7%: Score 33.8; DB 18; Length 580073;
FT Best Local Similarity 35.6%: Pred. No. 6.9;
FT Matches 65; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
FT
FT QY 171 tgatgatgagtgagacagatgttttataragatcagccgacagtggttaaatgaacc 230
FT      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
FT Db 494944 TGATGATGACACACAGTGTGTTTGTGATTCACAAAACATRTGGTGTAAAGAAAGAGC 494885
FT
FT QY 231 taactttactataaagtgtgtctcctcaggggtgaagctcctcgctcaggtttc 287
FT      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
FT Db 494884 TGCACATATATCTTAATGTGAATTTGCGTGTGTTTACCACACTGATTTCCGCTTT 494828
FT
FT RESULT 9
FT Q84832/c
FT ID Q84832 standard; DNA; 234 BP.
FT XX
FT AC Q84832;
FT XX
FT DT 25-SEP-1995 (first entry)
FT XX
FT DE Spino cerebellar ataxia type 1 CAG repeat region patient #2.
FT XX
FT KW Spino cerebellar ataxia type 1; SCA 1; presymptomatic diagnosis;
FT KM CAG repeat region; patient #2; ss.
FT XX

```

```

OS Homo sapiens.
XX
XX WO9501437-A.
XX
XX 12-JAN-1995.
XX
XX PD 29-JUN-1994; 94WO-US07336.
XX
XX PF 29-JUN-1994; 93US-0084365.
XX
XX PR 29-JUN-1993; 94US-0267803.
XX
XX PR 28-JUN-1994; 94US-0267803.
XX
XX PA (MINU ) UNIV MINNESOTA.
XX
XX PI Chung M, Orr HT, Zoghbi HY;
XX
XX DR WPI; 1995-061001/08.
XX
XX PT New autosomal dominant spino cerebellar ataxia type 1 nucleic acid
XX PT - used to develop prods. for detection or presymptomatic
XX PT diagnosis of a SCA1 disorder
XX
XX PS Disclosure; Fig 2; 11pp; English.
XX
XX CC Q84831-Q84835 show the CAG repeat regions of five individuals
XX CC affected with spino cerebellar ataxia type 1 (SCA 1). It is within
XX CC CAG repeat region (Q84804) that the mutations responsible for
XX CC SCA 1 occur. The full nucleic acid (Q84793) and its protein product
XX CC (R7111) can be used to develop products, for the presymptomatic
XX CC detection of a SCA 1 disorder.
XX
XX SQ Sequence 234 BP; 73 A; 80 C; 77 G; 4 T; 0 other;
XX
XX Query Match          10.6%: Score 33.6; DB 16; Length 234;
XX Best Local Similarity 47.2%: Pred. No. 0.24;
XX Matches 102; Conservative 0; Mismatches 114; Indels 0; Gaps 0;
XX
XX QY 95 gctcgtaccacatgcaggtgactagctacttagccacacatgcaatgactcaggtgc 154
XX      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
XX Db 216 GCTGAGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 157
XX
XX QY 155 ctactgacacgtgacttgatgatgagtgagacagatgtttttagatgatacgcgcagc 214
XX      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
XX Db 156 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 97
XX
XX QY 215 ggttaaatgcaaacacttactataaagtgtgtctcctcaggtgtaagctccg 274
XX      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
XX Db 96 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 37
XX
XX QY 275 gcttcaggtttttagctgtgtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 310
XX      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
XX Db 36 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1
XX
XX RESULT 10
XX Z56367
XX ID Z56367 standard; DNA; 1185 BP.
XX XX
XX AC Z56367;
XX XX
XX DT 17-MAR-2000 (first entry)
XX XX
XX DE Escherichia coli flagellin protein nucleotide sequence SRQ ID NO:37.
XX XX
XX KW Flagellin; fliC; antigen; detection; ds.
XX XX
XX OS Escherichia coli.
XX XX
XX PN WO9961458-A1.
XX
XX PD 02-DEC-1999.
XX
XX PF 21-MAY-1999; 99WO-AU00385.
XX

```

```

XX PR 21-MAY-1998; 98AU-0003634.
XX XX
XX (UNSY ) UNIV SYDNEY.
XX PA
XX PI Reeves PR, Wang L;
XX DR WPI; 2000-072598/06.
XX PT Novel nucleic acid molecule useful for the detection of flagellated
XX bacterial strains in food, faeces, etc.
XX PS Claim 3; Page 211-212; 245pp; English.
XX CC 256331 to 256398 represent nucleic acid molecules (1) encoding all or
XX part of an Escherichia coli flagellin protein except a protein
XX expressed by E. coli H1, H7, H12 or H48 type strains. The present
XX invention also describes a method of detecting the presence of E. coli
XX of a particular H serotype in a sample, comprising specifically
XX hybridising a nucleic acid, preferably at least a pair, derived from a
XX flagellating gene, specific for a particular flagellin gene associated
XX with the H serotype, to any E.coli in the sample which contain the gene,
XX and detecting any hybridised molecules, identifying the presence of that
XX serotype in the sample. (1) are useful for: (1) detecting the presence
XX of E. coli of H serotype in a sample by hybridising at least one or a
XX pair of (1) to any E. coli in the sample and detecting the hybridised
XX nucleic acid molecules; and (2) for detecting the presence of both O
XX and H-serotypes of E. coli by hybridising at least one or a pair of (1)
XX to any E. coli present in the sample and detecting the hybridised
XX nucleic acid molecules. (1) is particularly useful for detecting the
XX combination of O and H antigen. Hybridised (1) when using at least one
XX (1) is detected by southern blot analysis and, when using a pair of (1),
XX is detected by polymerase chain reaction (PCR). 256399 to 256420
XX represent primers used in the exemplification of the present invention.
XX SO Sequence 1185 BP; 294 A; 310 C; 292 G; 289 T; 0 other;

Query Match 10.6%; Score 33.6; DB 21; Length 1185;
Best Local Similarity 46.2%; Pred. No. 0.5;
Matches 111; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 43 gcaataatgctgctagaggtatagtgtaccatgccacaataacagagtgctctgt 102
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 733 gctctaatagtgtactacgycgcactcagaacggtcagctcagatcaagtaactctgac 792
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 103 accaatgaggtgacttgacttgaccacacatcaggtactcagctgctcctatgac 162
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 793 gccaaaggtgagctgtgtgttaacgttaacatcagggttaaaatactcagcgtgtgcg 852
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 163 actgcaacttgatgagtgacagatgtttttagatagatcagcgcacagtggttaaa 222
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 853 accgggtgtgacgttcctgagcgaacagcgggtgtgagctccaactaacagcgttgatacc 912
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 223 tgcgaacttaacttacttaataatggtgtctctccacaggtgtaagctcgtgcctcag 282
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 913 gttactctgcaactcagcagctgactggtgcaactactgacgcgaaggtactcgaactcag 972
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
X89891/c
ID X89891 standard; DNA; 397 BP.
XX
XX X89891;
XX AC
XX DT 05-NOV-1999 (first entry)
XX DE Spino cerebellar ataxia type III (SCAIII) gene fragment.
XX XX
XX Spino cerebellar ataxia type III; SCAIII; reverse dot hybridisation;
XX KM PCR-microplate hybridisation; PCR-MPH; trinucleotide repeat; TNR;
XX KM SCAIII syndrome; ss.
XX XX

```

```

OS Homo sapiens.
XX XX Location/Qualifiers
XX FH Key 137..355
XX FT repeat_region /*tag= a
XX FT repeat_unit 137..139
XX FT /*tag= b
XX FT /note= "trinucleotide repeat"
XX XX
XX PN WC0943852-A1.
XX PD 02-SEP-1999.
XX PE 18-FEB-1999; 99WO-KR00078.
XX PR 26-FEB-1998; 98KR-0006278.
XX PA (JIND/) JIN D K.
XX PI (SMSU ) SAMSUNG FINE CHEM CO LTD.
XX XX
XX JIN DK;
XX DR WPI; 1999-527634/44.
XX PT Diagnosis of spinocerebellar ataxia type III (SCA III) syndrome
XX using techniques which ensure highly accurate diagnosis
XX PS Claim 1; Page 12-13; 28pp; English.
XX CC The invention relates to the diagnosis of spinocerebellar ataxia type
XX III (SCAIII) syndrome using reverse dot hybridisation or PCR-microplate
XX hybridisation (PCR-MPH). The method comprises attaching a portion of the
XX SCAIII gene containing 73 copies of the trinucleotide (CAG) repeat unit
XX (the present sequence) to a substrate, and hybridising with amplified
XX testee genomic DNA containing copies of the trinucleotide units. PCR
XX amplified with labeled primers (X89889-90). The new method is useful for
XX diagnosis of SCAIII syndrome, and for determining the severity of the
XX disease. The present sequence represents the SCAIII gene fragment
XX containing 73 trinucleotide (TNR) repeats.
XX SO Sequence 397 BP; 124 A; 104 C; 104 G; 65 T; 0 other;

Query Match 10.6%; Score 33.4; DB 20; Length 397;
Best Local Similarity 47.0%; Pred. No. 0.36;
Matches 103; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 92 tagctctggtaccatgacaggtgacttagctactttagccacacatgactactcagt 151
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 364 TAGGTCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 305
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 152 gtccactgagcagctgacatgatgatgagatgacagatgttttgatagatcagccgac 211
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 304 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 245
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 212 agtgtgtaaatgcaaacactaacttactataatggtgtgtcctccacaggtgaagctc 271
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 244 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 185
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 272 ctgagctcaggttttctgctgctgctgctgctgctgctgctgctgctgctgag 310
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 184 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 146
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
T27627
ID T27627 standard; cDNA to mRNA; 2322 BP.
XX
XX T27627;
XX AC
XX DT 14-NOV-1996 (first entry)
XX DE Malic enzyme coding sequence #1.
XX XX

```

XX Maltic enzyme; Aloe; decarboxylase; malic acid; photosynthesis; plant;
 KW carbonic acid; Calvin cycle; ss.
 XX Aloe arborescens.
 OS
 XX
 FH Key Location/Qualifiers
 FT 229..1986
 FT CDS /*tag= a
 FT /product= malic enzyme
 XX
 PN JP08089250-A.
 PD 09-APR-1996.
 XX
 PF 21-SEP-1994; 94JP-0226159.
 XX
 PR 21-SEP-1994; 94JP-0226159.
 XX
 PA (MITK) MITSUI TOATSU CHEM INC.
 XX
 DR WP1; 1996-233343/24.
 DR P-PSDB; R96246.
 XX
 PT Maltic enzyme gene from Aloe plants - promotes decarboxylation from
 PT malic acid and ultimately confers ability to fix carbonic acid
 PT through Calvin cycle
 XX
 PS Claim 1; Page 5-8; 12pp; Japanese.
 XX
 SQ Sequence 2322 BP; 655 A; 484 C; 543 G; 640 T; 0 other;
 XX
 CC This sequence represents the coding sequence for malic enzyme. This
 CC sequence was isolated from Aloe arborescens. This sequence can be
 CC introduced into plants, such as Aloe, to increase the level of malic
 CC enzyme activity. Introduction of this gene into a plant promotes
 CC decarboxylation from malic acid, and confers a photosynthetic ability on
 CC plants, thereby increasing their ability to fix carbonic acid through the
 CC Calvin cycle.
 CC
 XX
 SQ
 Query Match 10.4%; Score 32.8; DB 17; Length 2322;
 Best Local Similarity 55.2%; Pred. No. 1.2;
 Matches 64; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
 QY 16 gaagcattgtaactgagccttcgcagcaataatgctgtagagttatgtgacca 75
 Db 1771 gaagctctgctcagcagcaggtgacagaagaattcgtcaatgagctgattatccacc 1830
 QY 76 tgcacaataaacagagtagctcgtgtaacaaatgcaagtgtagctacttagc 131
 Db 1831 ttacacataatcagaagaatctcgcacacattgtagcagtaatgtagctgacaaagc 1886
 RESULT 13
 Z40460
 ID 240460 standard; CDNA; 1987 BP.
 XX
 AC 240460;
 XX
 DT 15-FEB-2000 (first entry)
 XX
 DE Human gonadotropin receptor partial coding sequence #4.
 XX
 KW Human; gonadotropin receptor; screening; ligand; biomedical research;
 KW biochemical research; drug; hormone; reproductive tissue; infertility;
 KW contraception; ss.
 XX
 XX Homo sapiens.
 OS
 XX EP950711-A2.
 PN
 XX 20-OCT-1999.
 PD

XX 02-FEB-1999; 99EP-0200303.
 XX
 PF 06-FEB-1998; 98EP-0200357.
 PR 27-JUL-1998; 98EP-0202519.
 PR 24-SEP-1998; 98EP-0203213.
 XX
 XX (ALKU) AKZO NOBEL NV.
 PA
 PI Van Der Spek PJ, Helkoop JC;
 PI
 DR WP1; 1999-563673/48.
 DR P-PSDB; Y53574.
 XX
 PT New 7 transmembrane gonadotropin receptors, useful for screening for
 PT hormone analogs and drugs -
 PT
 XX
 PS Claim 2; Page 16-17; 38pp; English.
 XX
 CC Sequences 240457-240464 represent fragments of 3 novel human gonadotropin
 CC receptor genes. The novel gonadotropin receptors can be used to screen
 CC for ligands of the receptors. This screen may be used in biomedical and
 CC biochemical research to develop new drugs targeted to the gonadotropin
 CC receptors. For example, hormone analogs which activate or inhibit the
 CC function of the gonadotropin receptors or classical gonadotropin
 CC receptors may be detected. The altered expression or dysfunction of
 CC the gonadotropin receptors causes conditions in reproductive tissues.
 CC The ligands can be used for the treatment of infertility or for
 CC contraception.
 CC
 XX
 SQ Sequence 1987 BP; 353 A; 604 C; 564 G; 466 T; 0 other;
 XX
 CC
 Query Match 10.3%; Score 32.6; DB 20; Length 1987;
 Best Local Similarity 60.9%; Pred. No. 1.3;
 Matches 53; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
 QY 224 gcaaacctaactttacataatggtgtctccccaaggtgaagctcctgagcctcag 283
 Db 902 gcttcctcagcttcgctcgcctcagtgctgctcttcctcgtcacgcccagggcgtcaag 961
 QY 284 ttttgcctgctggtgctgcgcgtcag 310
 Db 962 tctgtcctgctggtgctgctgcctc 988
 RESULT 14
 T59196/C
 ID T59196 standard; CDNA; 2769 BP.
 XX
 AC T59196;
 XX
 DT 17-JUN-1997 (first entry)
 XX
 DE Neuronal alpha-bungarotoxin binding protein alpha1 subunit CDNA.
 XX
 DE Neuronal alpha-bungarotoxin binding protein alpha 1; cholinergic;
 KW ligand binding; ion channel; ss.
 KW
 XX
 OS Gallus sp.
 XX
 XX
 FH Key Location/Qualifiers
 FH 71..1513
 FT CDS /*tag= a
 FT sig_peptide 71..136
 FT /*tag= b
 FT mat_peptide 137..1510
 FT /*tag= c
 XX
 XX US5599709-A.
 PN
 XX 04-FEB-1997.
 PD

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:47:59 ; Search time 83.05 Seconds
(without alignments)
43.231 Million cell updates/sec

Title: US-09-196-161D-1

Perfect score: 538
Sequence: 1 GAAQGEANGNQPFANNAAR.....PQGEAPGVYFAAGAAAGV 105

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries.

Database :

A_Geneseq_36:*

1: /SIDSL/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseqp/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	84	15.1	613	19 W73009	Cobra venom protea
2	84	15.1	621	19 W73013	Cobra venom mocarh
3	77	13.8	424	17 W04321	Ancylostoma secret
4	75	13.4	1111	17 P91428	Kalinin/Laminin 5
5	75	13.4	1193	17 R91427	Kalinin/Laminin 5
6	74	13.4	3084	19 W50891	Mouse laminin A ch
7	74	13.3	3084	10 P94758	Sequence of mouse
8	72.5	13.0	686	18 W25719	Human alpha meltri
9	72.5	13.0	735	20 Y41695	Human PRO545 prote
10	72	12.9	410	20 W93577	Human AORP protei
11	72	12.9	468	19 W64483	Human DR4 protein.
12	72	12.9	468	20 Y31602	Human death recept

13	72	12.9	468	20 W93609	Human DR4 protein.
14	71	12.7	156	16 R70151	Amino terminal reg
15	71	12.7	1713	16 R70148	Deduced sequence o
16	70.5	12.6	320	19 W70220	Leishmania antigen
17	70.5	12.6	320	19 W70236	Leishmania antigen
18	70.5	12.6	1607	19 W50897	Mouse laminin G1 c
19	69	12.4	1086	21 Y84111	Amino acid sequenc
20	69	12.4	1086	21 Y84116	Amino acid sequenc
21	69	12.4	1086	21 Y58835	Corn cellulose syn
22	68.5	12.3	495	20 Y59972	Human endometrium
23	68.5	12.3	969	14 R41662	Paired basic amino
24	67	12.0	72	19 W68445	Australian paralys
25	67	12.0	144	18 W50286	Human Fas antigen
26	67	12.0	159	18 W50288	Human Fas antigen
27	67	12.0	314	16 R76238	Fas-delta-TM. Hom
28	67	12.0	314	17 R99682	Human Fas soluble
29	67	12.0	335	13 R28084	Human cell surface
30	67	12.0	335	16 R78606	Human Fas protein.
31	67	12.0	335	17 R99681	Human Fas antigen.
32	67	12.0	335	17 R92528	hFas from plasmid
33	67	12.0	335	18 W50289	Human Fas antigen.
34	67	12.0	335	19 W49104	Fas protein. Mamm
35	67	12.0	376	18 W50287	Human Fas antigen
36	67	12.0	376	19 W60037	Antigenic peptide
37	67	12.0	600	16 R78610	Expression vector
38	67	12.0	669	19 W64484	Human TNFRI protei
39	67	12.0	3075	19 W50892	Human laminin A ch
40	66.5	11.9	225	16 R88270	Papilloma virus ma
41	66	11.8	111	17 W01940	EGF-like repeats 3
42	66	11.8	314	20 W98070	Soluble Fas recept
43	66	11.7	1084	19 W33818	Arabidopsis cellu
44	65.5	11.6	102	18 W10370	Human Del-1 splice
45	65	11.6	1105	20 Y15459	SPO ID 5 of W09919

ALIGNMENTS

RESULT 1	W73009	standard; protein: 613 AA.
ID	W73009	
XX	W73009:	
AC	02-FEB-1999	(first entry)
XX		
DT		
XX		
DE	Cobra venom protease	mocarhagin NMW-9.
XX		
KW	Mocarhagin: snake venom; Mozambiquan spitting cobra; protease;	
KW	Inflammation; myocardial infarction; thrombosis; infection;	
KW	metastasis; therapy: NMW-9.	
XX		
OS	Naja mossambica	mossambica.
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..23
FT		/Label= Sig_peptide
FT	Protein	24..613
FT		/Label= Mat_protein
XX		
PN	W09846771-A2.	
XX		
PD	22-OCT-1998.	
XX		
PE	14-APR-1998;	98WO-US07998.
XX		
PR	18-FEB-1998;	98US-0026001.
PR	15-APR-1997;	97US-0843373.
PR	23-JAN-1998;	98US-0012637.
XX		
PA	(GENY) GENETICS INST INC.	
XX		
PI	Boodhoo A, Sako D, Seehra JS, Shaw G;	

XX WPI: 1998-568735/48.
 DR N-PSDB: V07897.
 OS
 XX
 PT Isolated mocarhagin cobra venom protease, and nucleic acids encoding
 PT it - used to develop products for treating e.g. myocardial
 PT infarction, thrombosis, bacterial or viral infection, metastatic
 PT conditions or inflammatory disorders
 XX
 PS Claim 40; Page 45-48; 97pp; English.

CC This is the amino acid sequence of mocarhagin NMW-9, a highly
 CC specific metalloproteinase from the venom of the Mozambiquan
 CC splitting cobra. The invention provides mocarhagin polypeptides
 CC (see W73007-13) and polynucleotides (see V07895-901) encoding them,
 CC as well as host cells and methods of producing the (especially
 CC mature) polypeptides. Mocarhagin proteins are capable of cleaving
 CC anionic polypeptide containing sulphated tyrosine residues,
 CC P-selectin glycoprotein (GP) ligand-1 (PSGL-1) and GPIIb- α IIb
 CC (claimed). They also inhibit neutrophil/HL60 binding, inhibit
 CC platelet binding to von Willebrand Factor, require Ca²⁺ and Zn²⁺
 CC ions for activity and have activity inhibited by excess EDTA or
 CC high concentrations of DFP (claimed). They can be used to inhibit
 CC selectin-mediated binding and to treat inflammatory disease
 CC (claimed). In particular, they can be used to treat e.g. myocardial
 CC infarction, vessel restenosis, thrombosis, bacterial or viral
 CC infection, metastatic conditions, inflammatory disorders such as
 CC arthritis, acute respiratory distress syndrome, asthma, emphysema,
 CC delayed type hypersensitivity reaction, systemic lupus
 CC erythematosus, thermal injury such as burns or frostbite,
 CC autoimmune thyroiditis, experimental allergic encephalomyelitis,
 CC multiple sclerosis, multiple organ injury syndrome secondary to
 CC trauma, diabetes, Reynaud's syndrome, neutrophilic dermatosis
 CC (Sweet's syndrome), inflammatory bowel disease, Grave's disease,
 CC glomerulonephritis, gingivitis, periodontitis, haemolytic uraemic
 CC syndrome, ulcerative colitis, Crohn's disease, necrotising
 CC enterocolitis, granulocyte transfusion associated syndrome,
 CC cytokine-induced enterocolitis, granulocyte transfusion associated
 CC syndrome, or cytokine-induced toxicity. Mocarhagin protein may
 CC also be useful in organ transplantation, both to prepare organs for
 CC transplantation and to quell organ transplant rejection, to treat
 CC haemodialysis and leukopheresis patients, or as an inhibitor of P-
 CC or E-selectin-mediated intercellular adhesion.

XX Sequence 613 AA:

SO Query Match 15.1%; Score 84; DB 19; Length 613;
 Best Local Similarity 29.2%; Pred. No. 0.4;
 Matches 28; Conservative 10; Mismatches 36; Indels 22; Gaps 5;

OY 11 QEPANNAARGLICVPCQINRVSVTNAG---DLATLATQCTOCPGTALDDGVDFD 66
 DB 443 qphagqd-segceckkfkfgagaacraakddcdpeltcggaacp-----ldltfq 492
 OY 67 RSAACVKKCKPMFYVNGSP-----QGEAPGVQY 95
 DB 493 tnglpc-qnegycyngkpcimlncialrpgvkv 527

RESULT 2
 W73013
 ID W73013 standard; Protein: 621 AA.
 AC W73013;
 DT 02-FEB-1999 (first entry)
 XX Cobra venom mocarhagin NMW-9ek.
 DE
 XX Mocarhagin; snake venom; Mozambiquan splitting cobra; protease;
 KW inflammation; myocardial infarction; thrombosis; infection;
 KM metastasis; therapy; NMW-9ek.

XX Naja mossambica mossambica.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..196
 FT Cleavage-site 192..196
 FT /label= Pro-peptide
 FT /note= "enterokinase cleavage site"
 FT Protein 197..621
 FT /label= Mat_protein

XX W09846771-A2.

XX 22-OCT-1998.

XX 14-APR-1998; 98WO-US07998.

XX 18-FEB-1998; 98US-0026001.

XX 15-APR-1997; 97US-0843373.

XX 23-JAN-1998; 98US-0012637.

XX (GENY) GENETICS INST INC.

XX Boodhoo A, Sako D, Seehra JS, Shaw G;

XX WPI: 1998-568735/48.

XX N-PSDB: V07901.

XX Claim 80; Page 65-68; 97pp; English.

XX This is the amino acid sequence of a modified cobra venom mocarhagin

XX protein, termed NMW-9ek, that includes an enterokinase cleavage site

XX between the propeptide and mature peptide of mocarhagin (see also

XX W73009). Introduction of the cleavage site may allow secretion of

XX active mocarhagin from eukaryotic host cells. The invention

XX provides mocarhagin polypeptides (see W73007-13) and polynucleotides

XX (see V07895-901), as well as host cells and methods of producing

XX of cleaving anionic polypeptides. Mocarhagin proteins are capable

XX residues, P-selectin glycoprotein (GP) ligand-1 (PSGL-1) and

XX GPIIb- α IIb (claimed). They also inhibit neutrophil/HL60 binding,

XX Zn²⁺ ions for activity and have activity inhibited by excess EDTA

XX or high concentrations of DFP (claimed). They can be used for

XX inhibiting selectin-mediated binding and for treating an

XX inflammatory disease (claimed). In particular, they can be used

XX for treating e.g. myocardial infarction, vessel restenosis,

XX thrombosis, bacterial or viral infection, metastatic conditions,

XX inflammatory disorders such as arthritis, acute respiratory distress

XX syndrome, asthma, emphysema, delayed type hypersensitivity reaction,

XX systemic lupus erythematosus, thermal injury e.g. burns or frostbite,

XX autoimmune thyroiditis, experimental allergic encephalomyelitis,

XX multiple sclerosis, multiple organ injury syndrome secondary to

XX trauma, diabetes, Reynaud's syndrome, neutrophilic dermatosis

XX (Sweet's syndrome), inflammatory bowel disease, Grave's disease,

XX glomerulonephritis, gingivitis, periodontitis, haemolytic uraemic

XX syndrome, ulcerative colitis, Crohn's disease, necrotising

XX enterocolitis, granulocyte transfusion associated syndrome,

XX cytokine-induced enterocolitis, granulocyte transfusion associated

XX syndrome, or cytokine-induced toxicity. Mocarhagin protein may

XX also be useful in organ transplantation, both to prepare organs for

XX transplantation and to quell organ transplant rejection, to treat

XX haemodialysis and leukopheresis patients, or as an inhibitor of P-
 or E-selectin-mediated intercellular adhesion.

XX Sequence 621 AA;


```

XX AC R91427;
XX 13-NOV-1996 (first entry)
XX DE Kalinin/laminin 5 gamma-2 chain.
XX KW Kalinin; laminin; epidermolysis bullosa; junctional; probe;
XX KW detection; inhibit; monitor; malignancy.
XX OS Homo sapiens.
XX PN W09610646-A1.
XX PD 11-APR-1996.
XX PF 04-OCT-1995; 95WO-EP03918.
XX PR 04-OCT-1994; 94US-0317450.
XX PA (TRYG/) TRYGVASON K.
XX PI Kalunki P, Pyke C, Trygvason K;
XX DR WPI, 1996-209366/21.
XX DR N-PSDB; T13323.
XX PT Detection of kalinin or laminin 5 expression in cells - useful to
XX PT detect, monitor and inhibit the invasive growth of cell in tissue,
XX PT partic. malignant tissue
XX PS Disclosure; Fig 4A: 37pp; English.
XX CC The present sequence is the kalinin/laminin 5 gamma-2 chain. The gamma-2
XX CC chain is of importance to patients suffering from epidermolysis bullosa,
XX CC esp. the junctional form (JEB). Probes and antisense gamma-2 sequences
XX CC derived from this sequence can be used to detect, monitor and inhibit
XX CC the invasive growth of cells in tissue, partic. malignant tissue.
XX SQ Sequence 1193 AA;

Query Match 13.4%; Score 75; DB 17; Length 1193;
Best Local Similarity 26.7%; Pred. No. 8.1;
Matches 35; Conservative 11; Mismatches 25; Indels 60; Gaps 9;

OY 20 RGCVCQCNBRGS-----VTNAG---DLATLAT 45
DB 78 rdtclpcnckskyslsarcdnsgrcckpvgarcdclpsflmlltdagctqtdqllds 137
OY 46 QCSSTOCPTGTALDDGVTVDFDRSAACVCKKP-----NEYYN--GGSPGGEAPG 92
DB 138 kcdcd-pagla---gpcd-----agrcv-ckpavgercdrcrsgyynldggnpeg----c 184
OY 93 VQVFAAGAAA 103
DB 185 tqcfcyghsas 195

RESULT 6
W50891
ID W50891 standard; Protein; 3084 AA.
XX W50891;
XX 07-DEC-1998 (first entry)
XX DE Mouse laminin A chain.
XX KW Laminin; mouse; beta-amyloid; amyloidosis; Alzheimer's disease;
XX KW Down's syndrome; hereditary cerebral haemorrhage; inflammation;
XX KW malignancy; Familial Mediterranean Fever; multiple myeloma;
XX KW type II diabetes; prion disease; Creutzfeldt-Jacob disease; CJD;

```

```

KW KW Gertsmann-Straussler syndrome; kuru; scrapie; haemodialysis;
KW KW carpal tunnel syndrome; senile cardiac amyloid polynuropathy;
KW KW Familial Amyloidotic Polynuropathy; thyroid carcinoma; diagnosis;
KW KW therapy.
XX OS Mus sp.
XX FH Key
XX FT Domain
XX FT Location/Qualifiers
XX FT 2746..2922
XX FT /note="fourth globular domain repeat (Claim 13)"
XX FT Region
XX FT 2690..2700
XX FT /note="beta-amyloid protein binding region
XX FT (Claim 12)"
XX PN W09815179-A1.
XX PD 16-APR-1998.
XX PF 08-OCT-1997; 97WO-US18145.
XX PR 08-OCT-1996; 96US-0027981.
XX PA (UNIV ) UNIV WASHINGTON.
XX PI Castlillo G, Snow AD;
XX DR WPI, 1998-240534/21.
XX PT Use of laminin and fragments - for developing products for use in
XX PT the diagnosis and treatment of amyloid disease, e.g. Alzheimer's
XX PT disease or CJD
XX PS Claim 15; Page 74-79; 132pp; English.
XX CC This is the amino acid sequence of the mouse laminin A chain. The
XX CC primary object of the invention is to use laminin, laminin-derived
XX CC protein fragments and/or laminin-derived polypeptides as potent
XX CC inhibitors of amyloid formation, deposition, accumulation and/or
XX CC persistence in Alzheimer's disease and other amyloidoses. The
XX CC laminin products (see W50888-98) may include mouse or human laminin
XX CC A or A1 chain, laminin B1 or B2 chain, laminin A2 chain (merostin),
XX CC the beta-amyloid binding domain of the laminin A1 chain and
XX CC method for treating an amyloid disease comprises administering a
XX CC polypeptide having a conformational similarity to a fragment of a
XX CC laminin protein. A method for diagnosing an amyloid disease
XX CC involves determining levels of laminin in a sample. Production
XX CC of laminin or its fourth globular repeat in vivo provides a method
XX CC for in vivo inhibition of beta-amyloid amyloidosis. The products
XX CC and methods can be used for the diagnosis, prognosis, monitoring
XX CC and treatment of amyloidoses such as Alzheimer's disease, Down's
XX CC syndrome and hereditary cerebral haemorrhage with amyloidosis of
XX CC the Dutch type (where the specific amyloid is the beta-amyloid
XX CC protein), the amyloidosis associated with chronic inflammation,
XX CC various forms of malignancy and Familial Mediterranean Fever (AF
XX CC amyloid or inflammation-association amyloidosis), the amyloidosis
XX CC associated with multiple myeloma and other B-cell abnormalities
XX CC (AL amyloid), the amyloidosis associated with type II diabetes
XX CC (amylin or islet amyloid), the amyloidosis associated with prion
XX CC diseases including Creutzfeldt-Jacob disease, Gertsmann-Straussler
XX CC syndrome, kuru and animal scrapie (PrP amyloid), the amyloidosis
XX CC associated with long-term haemodialysis and carpal tunnel syndrome
XX CC (beta 2-microglobulin amyloid), the amyloidosis associated with
XX CC senile cardiac amyloid and Familial Amyloidotic Polynuropathy
XX CC (prealbumin or transthyretin amyloid), and the amyloidosis
XX CC associated with endocrine tumours such as medullary carcinoma of
XX CC the thyroid (variant of procalcitonin).
XX SQ Sequence 3084 AA;

Query Match 13.4%; Score 75; DB 19; Length 3084;
Best Local Similarity 30.0%; Pred. No. 24;

```

	Matches	21;	Conservative	10;	Mismatches	31;	Indels	8;	Gaps	2;
OY	23	CVPCCINRVGSVTNMGADLTATLTCGTCCTCPTALDDCVTDVDFDSAAQCCKP-PNEY	81							
		:: :	:	::		:	:		: :	
Db	1094	cvpccinrvgsvtlptcdtlcgicscdsygcscskenvv-----gpcskcagttfal	1146							
OY	82	NGSPOGEAP	91							
		: :	:							
Db	1147	rghnpgqgcp	1156							
 RESULT 7										
ID	P94758	standard; protein: 3084 AA.								
XX										
AC	P94758;									
XX										
DT	27-JAN-1991	(first entry)								
XX										
DE	Sequence of mouse laminin A chain.									
XX										
KW	Peripheral nerve regeneration.									
XX										
OS	Mouse.									
XX										
FH	Key	Location/Qualifiers								
FT	Peptide	1..24								
FT	Protein	25..3084								
XX										
PN	U57267564-A.									
XX										
PD	28-FEB-1989.									
XX										
PE	07-NOV-1988;	88US-0267564.								
XX										
PR	07-NOV-1988;	88US-0267564.								
XX										
PA	(USSH) US DEPT HEALTH & HUMAN.									
XX										
PI	Yamada Y, Sasaki M, Kleiman HK, Martin GR;									
XX										
DR	WPI: 1989-138175/18.									
DR	N-PSDB: N91779.									
XX										
PT	DNA encoding human laminin A chain, used in vector system -									
PT	and new synthetic peptide(s) with laminin-type biological									
PT	activity									
XX										
PS	Disclosure: Figure 4; 90pp; English.									
XX										
CC	The sequence encoding mouse laminin A chain (N91779) is used as a probe									
CC	to screen a human cDNA library. Laminin is a very potent and rapid									
CC	stimulator of neurite outgrowth and promotes both central and peripheral									
CC	nerve regeneration.									
XX										
SQ	Sequence	3084 AA;								
 Query Match 13.3%; Score 74; DB 10; Length 3084; Best Local Similarity 28.6%; Pred. No. 31; Matches 20; Conservative 11; Mismatches 31; Indels 8; Gaps 2;										
OY	23	CVPCCINRVGSVTNMGADLTATLTCGTCCTCPTALDDGVTDVDFDSAAQCCKP-PNEY	81							
		:: :	:	::		:	:		: :	
Db	1094	cipccnltpclphcdlieglcsedsygcscskenvv-----ppqskcagttfal	1146							
OY	82	NGSPOGEAP	91							
		: :	:							
Db	1147	rghnpgqgcp	1156							
 RESULT 8										
425719										

[illegible]

secreted protein: transmembrane protein.

XX Homo sapiens.
 OS WO9946281-A2.
 PN 16-SEP-1999.
 PD
 XX 08-MAR-1999; 99MO-US05028.
 PF
 XX 10-MAR-1998; 98US-0077450.
 PR 11-MAR-1998; 98US-0077632.
 PR 11-MAR-1998; 98US-0077641.
 PR 11-MAR-1998; 98US-0077649.
 PR 12-MAR-1998; 98US-0077791.
 PR 13-MAR-1998; 98US-0078004.
 PR 17-MAR-1998; 98US-0040220.
 PR 20-MAR-1998; 98US-0078886.
 PR 20-MAR-1998; 98US-0078936.
 PR 20-MAR-1998; 98US-0078939.
 PR 25-MAR-1998; 98US-0079294.
 PR 26-MAR-1998; 98US-0079656.
 PR 27-MAR-1998; 98US-0079663.
 PR 27-MAR-1998; 98US-0079664.
 PR 27-MAR-1998; 98US-0079689.
 PR 27-MAR-1998; 98US-0079728.
 PR 27-MAR-1998; 98US-0079786.
 PR 30-MAR-1998; 98US-0079920.
 PR 30-MAR-1998; 98US-0079923.
 PR 31-MAR-1998; 98US-0080105.
 PR 31-MAR-1998; 98US-0080107.
 PR 31-MAR-1998; 98US-0080165.
 PR 31-MAR-1998; 98US-0080194.
 PR 01-APR-1998; 98US-0080327.
 PR 01-APR-1998; 98US-0080328.
 PR 01-APR-1998; 98US-0080333.
 PR 01-APR-1998; 98US-0080334.
 PR 08-APR-1998; 98US-0081049.
 PR 08-APR-1998; 98US-0081070.
 PR 08-APR-1998; 98US-0081071.
 PR 09-APR-1998; 98US-0081195.
 PR 09-APR-1998; 98US-0081203.
 PR 09-APR-1998; 98US-0081229.
 PR 15-APR-1998; 98US-0081817.
 PR 15-APR-1998; 98US-0081838.
 PR 15-APR-1998; 98US-0081952.
 PR 15-APR-1998; 98US-0081955.
 PR 21-APR-1998; 98US-0082568.
 PR 21-APR-1998; 98US-0082569.
 PR 22-APR-1998; 98US-0082700.
 PR 22-APR-1998; 98US-0082704.
 PR 22-APR-1998; 98US-0082804.
 PR 23-APR-1998; 98US-0082767.
 PR 23-APR-1998; 98US-0082796.
 PR 27-APR-1998; 98US-0083336.
 PR 28-APR-1998; 98US-0083322.
 PR 29-APR-1998; 98US-0083392.
 PR 29-APR-1998; 98US-0083495.
 PR 29-APR-1998; 98US-0083496.
 PR 29-APR-1998; 98US-0083499.
 PR 29-APR-1998; 98US-0083500.
 PR 29-APR-1998; 98US-0083545.
 PR 29-APR-1998; 98US-0083554.
 PR 29-APR-1998; 98US-0083558.
 PR 29-APR-1998; 98US-0083559.
 PR 30-APR-1998; 98US-0083742.
 PR 05-MAY-1998; 98US-0084366.
 PR 06-MAY-1998; 98US-0084414.
 PR 06-MAY-1998; 98US-0084441.
 PR 07-MAY-1998; 98US-0084598.
 PR 07-MAY-1998; 98US-0084600.
 PR 07-MAY-1998; 98US-0084627.

PR 07-MAY-1998; 98US-0084637.
 PR 07-MAY-1998; 98US-0084639.
 PR 07-MAY-1998; 98US-0084640.
 PR 07-MAY-1998; 98US-0084643.
 PR 13-MAY-1998; 98US-0085323.
 PR 13-MAY-1998; 98US-0085338.
 PR 13-MAY-1998; 98US-0085339.
 PR 15-MAY-1998; 98US-0085353.
 PR 15-MAY-1998; 98US-0085373.
 PR 15-MAY-1998; 98US-0085579.
 PR 15-MAY-1998; 98US-0085580.
 PR 15-MAY-1998; 98US-0085582.
 PR 15-MAY-1998; 98US-0085689.
 PR 15-MAY-1998; 98US-0085697.
 PR 15-MAY-1998; 98US-0085700.
 PR 15-MAY-1998; 98US-0085704.
 PR 18-MAY-1998; 98US-0086023.
 PR 22-MAY-1998; 98US-0086392.
 PR 22-MAY-1998; 98US-0086414.
 PR 22-MAY-1998; 98US-0086430.
 PR 22-MAY-1998; 98US-0086486.
 PR 28-MAY-1998; 98US-0087098.
 PR 28-MAY-1998; 98US-0087106.
 PR 28-MAY-1998; 98US-0087208.
 PR 30-JUL-1998; 98US-0094651.
 PR 11-SEP-1998; 98US-0100038.

XX (GETH) GENENTECH INC.

PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;

XX WPI; 1999-551358/46.

DR N-PSDB; Z33953.

PT New secreted and transmembrane polypeptides and their polynucleotides,
 PT useful for treating blood coagulation disorders, cancers and cellular
 PT adhesion disorders -

XX Claim 12; Fig 30; 530pp; English.

XX The present invention describes secreted and transmembrane polypeptides
 CC and their polynucleotides. The nucleotide sequences are useful as
 CC sources of probes, primers, for chromosome mapping, and for generation
 CC of antisense sequences. They can also be used to create transgenic
 CC animals. The proteins can be used to treat a variety of diseases and
 CC disorders, depending on their function. Diseases that may be treated
 CC include blood coagulation disorders, cancers and cellular adhesion
 CC disorders. They may also be used to raise antibodies. Z33891 to
 CC Z34338, and Y41685 to Y41774 represent polynucleotide and polypeptide
 CC sequence given in the exemplification of the present invention.

XX Sequence 735 AA;

Query Match 13.0%; Score 72.5; DB 20; Length 735;

Best Local Similarity 27.2%; Pred. No. 8.6;

Matches 25; Conservative 10; Mismatches 38; Indels 19; Gaps 4;

OY 19 ARGICV-PCOINRVGSV---TNAGDLATLATQCSTOCPTGTLDDGVDPVFRSAACV 73

Db 462 angiccedcqlkpagtaacdsnsclpelfcqaashcpaanvylhghs----- 510

OY 74 KCK--PNFYNGSGSPQGEAPGVGVFAAANA 103

Db 511 -cqvdygcynglclgchegcvrlwpgpkpa 541

RESULT 10

W93577

ID W93577 standard; Protein; 410 AA.

XX W93577;

DT 18-JUN-1999 (first entry)

XX DE Human APOB protein. .
 XX XX
 KW Tumour necrosis factor receptor; signal transducer molecule; TNF; APO4;
 KW developmental abnormality; gestational abnormality; prostate cancer;
 KW APO8; APO9; TNRL-1; TNRL-3; diagnosis; treatment; therapy; disease;
 KW cytoplasmic domain; immunogen; antibody preparation; breast carcinoma;
 KW apoptosis; human; APOB; APO-related protein.
 XX OS
 XX Homo sapiens.
 XX PN
 XX WO9911791-A2.
 XX PD
 XX 11-MAR-1999.
 XX PF
 XX 04-SEP-1998; 98WO-US18393.
 XX PR
 XX 05-SEP-1997; 97US-0924634.
 XX PA
 XX (UNITV) UNIV WASHINGTON.
 XX PI
 XX Chaudhary PM:
 XX DR
 XX WPI: 1999-205191/17.
 XX N-PSDB; X23411.
 XX PT
 XX New Tumor Necrosis Factor family receptor polypeptides and ligands -
 XX PT useful for diagnosis and treatment of prostate cancer and
 XX PT developmental or gestational abnormalities
 XX PS
 XX Example 1; Fig 3; 156pp; English.
 CC This invention describes isolated Tumor Necrosis Factor (TNF) family
 CC receptor polypeptides: APO4, APO6, APO8 and APO9 or their active
 CC fragments, and isolated TNF related ligands 1 and 3 (TNRL1 and TNRL3) or
 CC their active fragments. APO4 is useful for diagnosing prostate cancer
 CC by determining levels of APO4 in an individual. Prostate cancer can also
 CC be treated using APO4 selective binding agents linked to a therapeutic
 CC moiety. APO4 polypeptides are also useful for identifying selective
 CC binding agents, useful in diagnosis/treatment of disease by binding of
 CC agents to the polypeptide/active fragment which is extracellular, or
 CC expressed on the cell surface. The binding is preferably performed in
 CC vivo. APO4 polypeptides/active fragments are also useful for screening
 CC for agonists and antagonists by binding and observing the change in APO4
 CC activity. Effective pharmacological agents useful in diagnosis or
 CC treatment of disease are also identified using APO4 polypeptides/active
 CC fragments and APO4 signal transducer molecules that specifically interact
 CC with a cytoplasmic domain of APO4 and detecting a change in level of APO4
 CC activity. The method is performed in vivo or in vitro. APO polypeptides
 CC are all useful as immunogens for preparing antibodies. APO4 is also
 CC useful for diagnosis/treatment of developmental or gestational
 CC abnormalities. APO8 was transfected to human breast carcinoma cell line
 CC MCF-7, and induced apoptosis.
 CC XX
 CC Sequence 410 AA:
 SQ
 Query Match 12.9%; Score 72; DB 20; Length 410;
 Best Local Similarity 25.3%; Pred. No. 4.9; Indels 26; Gaps 4;
 Matches 24; Conservative 8; Mismatches 37;
 Oy 5 GEANGNOPPANNMARGI-----CVPQINRGSVTNAGDLATLQCTGCTPT 53
 | | | | | : | | | | | : | | | | | : | | | | |
 DB 77 gshserpgacnrcteggytynasnmfclpctackdeerspottltnlac--gckp 134
 | | | | | : | | | | | : | | | | | : | | | | |
 Oy 54 GTALDDGVTDFVDRSAQCCKPKPFYNGSGPQG 88
 | | | | | : | | | | | : | | | | | : | | | | |
 DB 135 glfrn-----dnasemcrkc-----stgcpgr 156
 | | | | | : | | | | | : | | | | | : | | | | |
 RESULT 11
 W64483
 ID W64483 standard; Protein: 468 AA.

XX AC W64483;
 XX XX
 DT 20-OCT-1998 (first entry)
 XX XX
 DE Human DR4 protein.
 KW Death domain containing receptor 4; DR4; apoptosis; cancer; inflammation;
 KW agonist; tumour necrosis factor; TNF; ligand; autoimmune disease;
 KW infection; graft rejection; antagonist; inhibitor; diagnostic.
 XX OS
 XX Homo sapiens.
 XX PN
 XX Key Location/Qualifiers
 XX FH 1..23
 XX FH /label= signal
 XX FT 24..468
 XX FT /label= DR4
 XX FT 24..238
 XX FT /label= extracellular_domain
 XX FT 239..264
 XX FT /label= transmembrane_domain
 XX FT 265..468
 XX FT /label= intracellular_domain
 XX FT 379..422
 XX FT /label= death_domain
 XX PN
 XX WO9832856-A1;
 XX PD
 XX 30-JUL-1998.
 XX PF
 XX 27-JAN-1998; 98WO-US01464.
 XX PR
 XX 05-FEB-1997; 97US-0037829.
 XX PR 28-JAN-1997; 97US-0035722.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PA (UNMI) UNIV MICHIGAN.
 XX PI Dixit VM, Gentz RL, Ni J, Pan JG, Rosen CA;
 XX DR WPI: 1998-427952/36.
 XX N-PSDB; V49527.
 XX PT
 XX Nucleic acid encoding human death domain-containing receptor 4 -
 XX PT useful for therapeutic modulation of apoptosis, in e.g. cancer and
 XX PT autoimmune diseases
 XX PS
 PS Claim 1a; Fig 1; 92pp; English.
 CC This sequence represents a human death domain containing receptor 4, DR4.
 CC DR4 agonists are used to increase apoptosis induced by tumour necrosis
 CC factor (TNF)-family ligands, e.g. in cases of cancer, autoimmune disease,
 CC viral or other infections, inflammation, graft vs. host disease, acute or
 CC chronic graft rejection. Antagonists of DR4 are used to inhibit such
 CC apoptosis, e.g. in cases of acquired immune deficiency syndrome,
 CC neurodegenerative disease, myelodysplastic syndrome, ischaemic injury,
 CC toxin-induced liver damage, septic shock, cachexia and anorexia, also a
 CC wide range of inflammatory conditions. DR4 of fragments of the protein
 CC are used diagnostically, e.g. to detect mutant forms of DR4 (possibly
 CC associated with disease), for isolating the DR4 gene or related sequences
 CC and for chromosomal mapping.
 CC XX
 CC Sequence 468 AA:
 SQ
 Query Match 12.9%; Score 72; DB 19; Length 468;
 Best Local Similarity 25.3%; Pred. No. 5.7;
 Matches 24; Conservative 8; Mismatches 37; Indels 26; Gaps 4;
 Oy 5 GEANGNOPPANNMARGI-----CVPQINRGSVTNAGDLATLQCTGCTPT 53
 | | | | | : | | | | | : | | | | | : | | | | |
 DB 135 gshserpgacnrcteggytynasnmfclpctackdeerspottltnlac--gckp 192
 | | | | | : | | | | | : | | | | | : | | | | |


```

XX DE Amino terminal region of E170 encoded by alpha-3EPB and alpha-3EPB
XX DE cDNAs.
XX KM E170; epithelial ligand glycoprotein; epiligrin complex;
XX KM epithelial cell binding; alpha-3EPA.
XX OS Homo sapiens.
XX PN WO9506660-A.
XX PD 09-MAR-1995.
XX PF 02-SEP-1994; 94WO-US10261.
XX PR 02-SEP-1993; 93US-0115918.
XX PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX PI Carter WG, Gil SG, Ryan MC;
XX DR WPI: 1995-115398/15.
XX DR N-PSDB; Q83239.
XX PT New nucleic acid encoding epiligrin, an epithelial ligand complex
XX PT - also related vectors, transformed cells, proteins and
XX PT antibodies, useful therapeutically and diagnostically, e.g., in
XX PT cases of inflammation and to induce cancer cell differentiation.
XX PS Example: Fig 18C; 187pp; English.
XX CC Q83237/R70149, Q83238/R70150 and Q83239/R70151 illustrate the
XX CC sequence variability in domain IITA, near the amino-terminal
XX CC portion of the protein encoded by alpha-3EP. Multiple sequence
XX CC alignments indicate that there are two distinct alpha-3 transcripts
XX CC that display variability within domain IITA. The transcript
XX CC referred to as alpha-3EPB maintains homology to alpha-1 laminin
XX CC throughout domain IIT1 and into domain IV (see Q83238). cDNA clones
XX CC 5-4-1, 5-4-2 and 3-1-1 contain sequences which are absolutely
XX CC identical throughout the 3' end of each clone (Q83239); the most
XX CC reasonable explanation for which is that they represent two
XX CC different products of the same gene. The sequence provided in
XX CC Q83238 was deduced from sequencing cDNA clone 5-4-1 and does not
XX CC corresp. to the 5' end of the alpha-3EPB transcript.
XX SQ Sequence 156 AA:

Query Match 12.7%; Score 71; DB 16; Length 156;
Best Local Similarity 30.0%; Pred. No. 2.1;
Matches 24; Conservative 6; Mismatches 22; Indels 28; Gaps 6;

OY 17 NAARGIC--VPCQINR---VGSVTNAGDLATLATGCTGCTGALDDGVTDFDSAAQ 71
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 60 navhgsrscpcphntsfatgcvvngdv-----rcs--ckag-----ytlgtq 100
OY 72 CVKCKPNEYVN---GSPQ 87
   | : | | | : | | | |
DB 101 cercapgyfgnpqkfgyscq 120

RESULT 15
R70148
ID R70148 standard; Protein: 1713 AA.
XX
AC R70148;
XX
DT 05-OCT-1995 (first entry)
XX
DE Deduced sequence of cDNA corresp. to the alpha-3EPA transcript.
XX
KM E170; epithelial ligand glycoprotein; epiligrin complex;
KW epithelial cell binding.

```

```

XX OS Homo sapiens.
XX PN WO9506660-A.
XX PD 09-MAR-1995.
XX PF 02-SEP-1994; 94WO-US10261.
XX PR 02-SEP-1993; 93US-0115918.
XX PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX PI Carter WG, Gil SG, Ryan MC;
XX DR WPI: 1995-115398/15.
XX DR N-PSDB; Q83236.
XX PT New nucleic acid encoding epiligrin, an epithelial ligand complex
XX PT - also related vectors, transformed cells, proteins and
XX PT antibodies, useful therapeutically and diagnostically, e.g., in
XX PT cases of inflammation and to induce cancer cell differentiation.
XX PS Claim 12; Fig 15A-F; 187pp; English.
XX CC Q83235 depicts the nt. sequence compiled from sequencing cDNA
XX CC clones corresp. to the alpha-3 EPA transcript. The invention
XX CC includes nt sequences in the gp. comprising the nt sequence shown
XX CC in Q83235, the cDNA clone EP-1 (ATCC No. 75540) shown in Q83234,
XX CC the cDNA clone I-1 (ATCC No. 75539), and the cDNA clone 8-6 (ATCC
XX CC No. 75538), or the nt. sequences shown in Q83236. The entire nt
XX CC region encoding E170 is depicted in Q83236, and corresp. to the SQ
XX CC of alpha-3. Q83236 consists of a composite sequence derived from
XX CC several overlapping clones. A synthetic polypeptide of at least
XX CC 5 AAs that corresp. to part or all of the nt. sequence shown in
XX CC Q83236 is claimed.
XX SQ Sequence 1713 AA:

Query Match 12.7%; Score 71; DB 16; Length 1713;
Best Local Similarity 30.0%; Pred. No. 33;
Matches 24; Conservative 6; Mismatches 22; Indels 28; Gaps 6;

OY 17 NAARGIC--VPCQINR---VGSVTNAGDLATLATGCTGCTGALDDGVTDFDSAAQ 71
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 105 navhgsrscpcphntsfatgcvvngdv-----rcs--ckag-----ytlgtq 145
OY 72 CVKCKPNEYVN---GSPQ 87
   | : | | | : | | | |
DB 146 cercapgyfgnpqkfgyscq 165

Search completed: March 6, 2001, 12:49:30
Job time: 91 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:48:00 ; Search time 57.76 Seconds
(without alignments)
32.644 Million cell updates/sec

Title: US-09-196-161D-1
Perfect score: 558
Sequence: 1 GAAQGEANGNQPFANNNAAR.....PQGEAPGVYFAAGAAAGV 105

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_Aa: *
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6_COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/Backfilest.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	77	13.8	314	2	US-08-460-309-19
2	77	13.8	314	2	US-08-125-077-19
3	77	13.8	424	1	US-08-419-414-2
4	75	13.4	169	2	US-08-460-309-20
5	75	13.4	169	2	US-08-125-077-20
6	75	13.4	1111	1	US-08-317-4508-15
7	75	13.4	1111	3	US-08-800-593-15
8	75	13.4	1193	1	US-08-317-4508-13
9	75	13.4	1193	3	US-08-800-593-13
10	71	12.7	156	4	US-08-600-982-30
11	71	12.7	156	4	PCT-US94-10261A-30
12	71	12.7	1713	3	US-08-600-982-24
13	71	12.7	1713	4	PCT-US94-10261A-24
14	70.5	12.6	219	1	US-08-152-019A-32
15	68.5	12.3	969	2	US-08-284-941-2
16	68.5	12.3	969	2	US-08-447-642-2
17	68.5	12.3	969	4	PCT-US93-02147A-2
18	68	12.2	335	4	US-08-815-469-6
19	67	12.0	119	2	US-08-219-237B-3
20	67	12.0	219	3	US-08-974-022-45
21	67	12.0	314	1	US-08-444-231-19
22	67	12.0	314	1	US-08-152-443A-19
23	67	12.0	314	4	PCT-US95-17083-4
24	67	12.0	335	2	US-08-219-237B-2
25	67	12.0	335	2	US-08-409-338-1
26	67	12.0	335	4	PCT-US95-17083-2
27	67	12.0	3075	2	US-08-460-309-5
28	67	12.0	3075	2	US-08-125-077-5

29	66.5	11.9	341	2	US-08-209-521-11	Sequence 11, Appl
30	66.5	11.9	846	2	US-07-728-215-33	Sequence 33, Appl
31	66	11.8	111	1	US-08-288-728-4	Sequence 4, Appl
32	64.5	11.6	197	2	US-08-505-606-1	Sequence 1, Appl
33	64.5	11.6	801	1	US-07-906-349A-6	Sequence 6, Appl
34	63	11.3	610	1	US-08-365-470-3	Sequence 3, Appl
35	63	11.3	610	3	US-09-209-668-19	Sequence 19, Appl
36	63	11.3	610	5	5217870-2	Patent No. 5217870
37	63	11.3	1171	1	US-08-445-135-1	Sequence 1, Appl
38	63	11.3	3111	2	US-08-460-309-4	Sequence 4, Appl
39	63	11.3	3111	2	US-08-125-077-4	Sequence 4, Appl
40	62.5	11.2	225	1	US-08-152-019A-33	Sequence 33, Appl
41	62	11.1	693	2	US-08-380-403A-2	Sequence 2, Appl
42	62	11.1	693	2	US-08-380-403A-5	Sequence 5, Appl
43	62	11.1	693	2	US-08-895-628-2	Sequence 2, Appl
44	62	11.1	693	2	US-08-895-628-5	Sequence 5, Appl
45	62	11.1	803	3	US-09-063-035-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-460-309-19
Sequence 19, Application US/08460309
Patent No. 5837496
GENERAL INFORMATION:
APPLICANT: Engrvall, Eva
APPLICANT: Leivo, Ilmo
TITLE OF INVENTION: Nucleic Acids Encoding Metrosin, Metrosin
TITLE OF INVENTION: Fragments and Uses Thereof
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,309
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/125,077
FILING DATE: 22-SEP-1993
APPLICATION NUMBER: US PCT/US 94/10730
FILING DATE: 21-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/472,319
FILING DATE: 30-JAN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/919,951
FILING DATE: 27-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 9721
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 314 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-460-309-19


```
APPLICANT: Leivo, Ilmo
TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
TITLE OF INVENTION: Fragments and Uses Thereof
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSER: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,309
CLASSIFICATION: 435
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/125,077
FILING DATE: 22-SEP-1993
APPLICATION NUMBER: US PCT/US 94/10730
FILING DATE: 21-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/472,319
FILING DATE: 30-JAN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/919,951
FILING DATE: 27-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 9721
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 169 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-460-309-20

Query Match      13.4%; Score 75; DB 2; Length 169;
Best Local Similarity 26.7%; Pred. No. 0.5;
Matches 35; Conservative 11; Mismatches 25; Indels 60; Gaps 9;

QY 20 RGICVPCQINRVGS-----VTNAG---DLATLAT 45
DB 51 RDRCLPCNCNSKGSLSARCDNSGRCSCKPGVTGARCDRLCPGFHMLTDAGCTODRLIDS 110
QY 46 QCSFTGPTGTALDDGVTVDFDRSAACVCKRP-----NFYYN--GGSPGGEAPG 92
DB 111 KCDDC-PAGIA---GPCD-----AGRCV-CKPAVTGERCDRCRSRYNLDGSGNPEG---C 157
QY 93 VQVFAAGAAA 103
DB 158 TQCFYCHSAS 168

RESULT 5
US-08-125-077-20
Sequence 20, Application US/08125077
Patent No. 5872231
Patent No. 5872231 5840863
GENERAL INFORMATION:
APPLICANT: Engvall, Eva
APPLICANT: Leivo, Ilmo
TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
TITLE OF INVENTION: Fragments and Uses Thereof
```

```
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSER: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/125,077
FILING DATE: 22-SEP-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US 94/10730
FILING DATE: 21-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/472,319
FILING DATE: 30-JAN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/919,951
FILING DATE: 27-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 9721
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 169 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-125-077-20

Query Match      13.4%; Score 75; DB 2; Length 169;
Best Local Similarity 26.7%; Pred. No. 0.5;
Matches 35; Conservative 11; Mismatches 25; Indels 60; Gaps 9;

QY 20 RGICVPCQINRVGS-----VTNAG---DLATLAT 45
DB 51 RDRCLPCNCNSKGSLSARCDNSGRCSCKPGVTGARCDRLCPGFHMLTDAGCTODRLIDS 110
QY 46 QCSFTGPTGTALDDGVTVDFDRSAACVCKRP-----NFYYN--GGSPGGEAPG 92
DB 111 KCDDC-PAGIA---GPCD-----AGRCV-CKPAVTGERCDRCRSRYNLDGSGNPEG---C 157
QY 93 VQVFAAGAAA 103
DB 158 TQCFYCHSAS 168

RESULT 6
US-08-317-450B-15.
Sequence 15, Application US/08317450B
Patent No. 5660982
GENERAL INFORMATION:
APPLICANT: Tryggvason, Karl
APPLICANT: Kallunki, Pekka
APPLICANT: Pyke, Charles
TITLE OF INVENTION: Laminin Chains: Diagnostic and
TITLE OF INVENTION: Therapeutic Use
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSER: BANNER & ALLEGRETTI, LTD.
STREET: Ten South Wacker Drive
CITY: Chicago
```


Sequence 30, Application PC/TUS9410261A
GENERAL INFORMATION:
APPLICANT: Carter, William G.
APPLICANT: Gil, Susanna A.
APPLICANT: Ryan, Maureen C.
TITLE OF INVENTION: Epiligrin, an Epithelial Ligand for Integrins
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen, O'Connor, Johnson, and Kindness
STREET: 1420 Fifth Avenue
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-8100
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/10261A
FILING DATE: 02-SEP-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Shelton, Dennis K.
REGISTRATION NUMBER: 26,997
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682-8100
TELEFAX: (206) 224-0779
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: Amino terminal region of E170 encoded by the
DESCRIPTION: sequences shown in FIGURE 18C.
PCT-US94-10261A-30

Query Match 12.7% Score 71; DB 4; Length 156;
Best Local Similarity 30.0%; Pred. No. 1.2; Mismatches 22; Indels 28; Gaps 6;
Matches 24; Conservative 6; Mismatches 22; Indels 28; Gaps 6;
OY 17 NAARGIC--VPCQINR---VGSVTNAGDLATLTCSTQCPTGTALDDGVDFDPSAAQ 71
DB 60 NAVHSCRCACPCPHNPSFATGCVVNGDV-----RCS--CKAG-----YTGTQ 100
OY 72 CVKCKPNEYNN---GGSPO 87
DB 101 CERCAPGYFGNPKFGGSCQ 120
RESULT 12
US-08-600-982-24
Sequence 24, Application US/08600982
Patent No. 6120991
GENERAL INFORMATION:
APPLICANT: Carter, William G.
APPLICANT: Gil, Susanna A.
APPLICANT: Ryan, Maureen C.
TITLE OF INVENTION: Epiligrin, an Epithelial Ligand for Integrins
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen, O'Connor, Johnson, and Kindness
STREET: 1420 Fifth Avenue
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-8100
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/600,982,
FILING DATE: 02-SEP-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Shelton, Dennis K.
REGISTRATION NUMBER: 26,997
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682-8100
TELEFAX: (206) 224-0779
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 1713 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: E170 protein as translated from sequence
DESCRIPTION: of FIGURES 15A-15F, and as shown also in FIGURES
DESCRIPTION: 19A-19R
US-08-600-982-24

Query Match 12.7% Score 71; DB 3; Length 1713;
Best Local Similarity 30.0%; Pred. No. 22; Mismatches 22; Indels 28; Gaps 6;
Matches 24; Conservative 6; Mismatches 22; Indels 28; Gaps 6;
OY 17 NAARGIC--VPCQINR---VGSVTNAGDLATLTCSTQCPTGTALDDGVDFDPSAAQ 71
DB 105 NAVHSCRCACPCPHNPSFATGCVVNGDV-----RCS--CKAG-----YTGTQ 145
OY 72 CVKCKPNEYNN---GGSPO 87
DB 146 CERCAPGYFGNPKFGGSCQ 165

RESULT 13
PCT-US94-10261A-24
Sequence 24, Application PC/TUS9410261A
GENERAL INFORMATION:
APPLICANT: Carter, William G.
APPLICANT: Gil, Susanna A.
APPLICANT: Ryan, Maureen C.
TITLE OF INVENTION: Epiligrin, an Epithelial Ligand for Integrins
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen, O'Connor, Johnson, and Kindness
STREET: 1420 Fifth Avenue
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-8100
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/10261A
FILING DATE: 02-SEP-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Shelton, Dennis K.
REGISTRATION NUMBER: 26,997
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682-8100
TELEFAX: (206) 224-0779
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:

LENGTH: 1713 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 DESCRIPTION: E170 protein as translated from sequence of
 FIGURES 15A-15F, and as shown also in FIGURES 19A-19R
 PCT-US94-10261A-24

Query Match 12.7%: Score 71; DB 4; Length 1713;
 Best Local Similarity 30.0%: Pred. No. 22;

Matches 24: Conservative 6; Mismatches 22; Indels 28; Gaps 6;

OY 17 NARSGIC--VPCQINR---VGSVTNAGDLATLTCSTOCPTGTALDDGVTFDRSAO 71
 DB 105 NAVHSGRCRCPCPHINSFATGCVNGDV-----RCS--CKAG-----YTGQ 145
 OY 72 CYKCKPNFYNN---GGSFQ 87
 DB 146 CERCAPGYFGNPKFGGSCQ 165

RESULT 14

US-08-152-019A-32

; Sequence 32, Application US/08152019A

; Patent No. 5565331

; GENERAL INFORMATION:

; APPLICANT: Tessier-Lavigne, Marc

; APPLICANT: Serafini, Tito

; APPLICANT: Kennedy, Timothy

; APPLICANT: Placzek, Marysia

; APPLICANT: Jessell, Thomas

; TITLE OF INVENTION: NEURAL AXON OUTGROWTH MODULATORS

; NUMBER OF SEQUENCES: 46

; CORRESPONDENCE ADDRESS:

; ADDRESS: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT

; STREET: 4 Embarcadero Center, Suite 3400

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-4187

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentln Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/152,019A

; FILING DATE: 12-NOV-1993

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Osman, Richard Aron

; REGISTRATION NUMBER: 36,627

; REFERENCE/DOCKET NUMBER: A-59012/RAO

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 781-1989

; TELEFAX: (415) 398-3249

; TELEEX: 910 272299 FHT UR

; INFORMATION FOR SEQ ID NO: 32:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 219 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; US-08-152-019A-32

Query Match 12.6%: Score 70.5; DB 1; Length 219;
 Best Local Similarity 22.9%: Pred. No. 2.1;
 Matches 25; Conservative 9; Mismatches 26; Indels 49; Gaps 5;

OY 23 CVPCQINRVSVTNAGDLATLTCST-----OCPTGTALDD----- 59
 DB 110 CSRCHCSPIVGS-----LSYQCDISYGRCSCKRPVMDKCDRCQPGFHSILFEACRC 160
 OY 60 -----GVTVDFDRSAOCV-----KCKPNFY-YNCGSPGGEAP 91
 DB 161 SCDLRGSTDCEVNETGRVCVKDNVEGFNCERCKPGFHNLESSNPCKCTP 209

RESULT 15

US-08-284-941-2

; Sequence 2, Application US/08284941

; Patent No. 5863756

; GENERAL INFORMATION:

; APPLICANT: KIEFER, MICHAEL C

; APPLICANT: BARR, PHILIP J

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESS: COOLEY GODDARD CASTRO HUDDLESON & TATUM

; STREET: FIVE PALO ALTO SQUARE

; CITY: PALO ALTO

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94306

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentln Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/284,941

; FILING DATE: 2 August 1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: NEELEY PH.D., RICHARD L.

; REGISTRATION NUMBER: 30092

; REFERENCE/DOCKET NUMBER: CHIR-009/01US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 843-5070

; TELEFAX: (415) 857-0663

; TELEEX: 380816 COOLEY PA

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 969 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-284-941-2

Query Match 12.3%: Score 68.5; DB 2; Length 969;
 Best Local Similarity 25.7%: Pred. No. 21;
 Matches 18; Conservative 10; Mismatches 25; Indels 17; Gaps 2;

OY 23 CVPCQINRVSVTNAGDLATLTCSTOCPTGTALD-----DGVTVDFDRSAOC 72
 DB 712 CINCCHFSIGSVKT-----SRKCVSCPLGYFDDTARRCRCHKGCTCSSRAATOC 764
 OY 73 VKCKPNFYNN 82
 DB 765 LSCRRGFYHH 774

Search completed: March 6, 2001, 12:53:46
 Job time: 346 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:48:03 ; Search time 36.83 Seconds
(without alignments)
87.326 Million cell updates/sec

Title: US-09-196-161d-1

Perfect score: 558
Sequence: 1 GAAGGAGANGNPPANNNAAR.....PQGEAPGVYFAAGAAAGV 105

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	79	14.2	2569	1 LMA3_MOUSE	O61789 mus musculus
2	78	14.0	3635	1 LMA5_MOUSE	O61001 mus musculus
3	77	13.8	424	1 ASP_ANCCA	Q16937 ancylostoma
4	77	13.8	3712	1 LMA_DROME	Q00174 drosophila
5	75.5	13.5	713	1 TSA4_GIALA	P18499 giardia lam
6	75	13.4	1193	1 IMG2_HUMAN	Q13753 homo sapien
7	75	13.4	3084	1 LMA1_MOUSE	P19137 mus musculus
8	72.5	13.0	909	1 AD12_HUMAN	Q43184 homo sapien
9	71	12.7	1713	1 LMA3_HUMAN	Q16787 homo sapien
10	70.5	12.6	1607	1 LMG1_MOUSE	P02468 mus musculus
11	70	12.5	485	1 LEM2_BOVIN	P98107 bos taurus
12	69	12.4	1557	1 LML1_GAEEL	Q18823 caenorhabdi
13	68.5	12.3	415	1 TNRC_MOUSE	P50284 mus musculus
14	68.5	12.3	969	1 PAC4_HUMAN	P29122 homo sapien
15	68.5	12.3	1285	1 SL17_ENTHI	P23502 entamoeba h
16	68.5	12.3	1895	1 YLK3_GAEEL	P41951 caenorhabdi
17	68	12.2	1246	1 YMV2_GAEEL	P34504 caenorhabdi
18	67	12.0	335	1 FASA_HUMAN	P25445 homo sapien
19	67	12.0	1639	1 LMG1_DROME	P15215 drosophila
20	67	12.0	3075	1 LMA1_DROME	P25391 homo sapien
21	66.5	11.9	846	1 LMA1_HUMAN	P11554 drosophila
22	65	11.6	363	1 PGLR_ASPOB	P35335 aspergillus
23	65	11.6	541	1 YKCS_GAEEL	P41996 caenorhabdi
24	64.5	11.6	435	1 TNRC_HUMAN	P36941 homo sapien
25	64.5	11.6	867	1 SSPO_BOVIN	P88167 bos taurus
26	64.5	11.6	1955	1 AGR1_CHICK	P31656 gallus galli
27	64.5	11.6	2769	1 THYG_BOVIN	P01267 bos taurus
28	64.5	11.6	4543	1 LRP1_CHICK	P88167 gallus galli
29	64	11.5	328	1 C170_GIALA	P35799 giardia lam
30	64	11.5	444	1 DHE4_PPRRU	P45544 prevotella
31	64	11.5	2264	1 POL1_THRVS	P18522 tomato blac
32	64	11.5	3106	1 LMA2_MOUSE	Q60675 mus musculus
33	63.5	11.4	204	1 IP22_CAPAN	O49146 capsicum an

34	63.5	11.4	370	1 HMC2_DESVH	P33389 desulfiovibr
35	63.5	11.4	937	1 PAC4_RAT	Q63415 rattus norv
36	63	11.3	229	1 VG07_BP22	O01074 bacterioph
37	63	11.3	610	1 LEM2_HUMAN	P16581 homo sapien
38	63	11.3	764	1 HTR2_HAUSA	P17410 halobacteri
39	63	11.3	811	1 ES22_DROME	P34083 drosophila
40	63	11.3	873	1 ES21_DROME	P34082 drosophila
41	63	11.3	3110	1 LMA2_HUMAN	P24043 homo sapien
42	62.5	11.2	123	1 NLTP_PINTA	O41073 pinus taeda
43	62.5	11.2	316	1 CH12_SOLTU	P52404 solanum tub
44	62	11.1	445	1 DHE4_BACPR	P94316 bacteroides
45	62	11.1	571	1 DISJ_BOTTA	P30431 bothriops ja

ALIGNMENTS

RESULT 1
LMA3_MOUSE STANDARD: PRT: 2569 AA.
AC O61789: O61788: O61966;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 36, Last annotation update)
DE LAMININ ALPHA-3 CHAIN PRECURSOR (FRAGMENT).
GN LAMA3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
LN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE=LUNG;
RX MEDLINE=95394948; PubMed=7665604;
RA Galliano M.-F., Aberdam D., Aguzzi A., Ortonne J.-P., Meneguizzi G.;
RT "Cloning and complete primary structure of the mouse laminin alpha 3
RT chain. Distinct expression pattern of the laminin alpha 3A and alpha
RT 3B chain isoforms.";
RL J. Biol. Chem. 270:21820-21826(1995).
RN [2]
RP REVISIONS.
RA Aberdam D.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
LN [3]
RP SEQUENCE OF 1052-1770 FROM N.A.
RC TISSUE=LUNG;
RX MEDLINE=94281750; PubMed=8012114;
RA Aberdam D., Galliano M.-F., Mattei M.-G., Pisani-Spadafora A.,
RT Ortonne J.-P., Meneguizzi G.;
RL "Assignment of mouse nicein genes to chromosomes 1 and 18.";
RN Mamm. Genome 5:229-233(1994).
LN [4]
RP SEQUENCE OF 1052-1770 FROM N.A.
RC TISSUE=LUNG;
RX MEDLINE=94363405; PubMed=8081888;
RA Aberdam D., Aguzzi A., Baudoin C., Galliano M.-F., Ortonne J.-P.,
RL Meneguizzi G.;
RT "Developmental expression of nicein adhesion protein (laminin-5)
RT subunits suggests multiple morphogenic roles.";
RL Cell Adhes. Commun. 2:115-129(1994).
LN [5]
RP FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR. LAMININ
IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
-1- FUNCTION: LAMININ-5 IS THOUGHT TO BE INVOLVED IN (1) CELL ADHESION
VIA INTEGRIN ALPHA3/BETA-1 IN FOCAL ADHESION AND INTEGRIN ALPHA-
6/BETA-4 IN HEMIDESMOSOMES, (2) SIGNAL TRANSDUCTION VIA TYROSINE
PHOSPHORYLATION OF P125-FAK AND P80, (3) DIFFERENTIATION OF
KERATINOCYTES (BY SIMILARITY).
-1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
COMPRISED ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
THE ALPHA-3 CHAIN IS A SUBUNIT OF LAMININ-5 (EPILGRIN/KALININ/

CC NICEIN), AND POSSIBLY ALSO A COMPONENT OF LAMININ-6 (K-LAMININ)
 CC AND LAMININ-7 (KS-LAMININ).
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR, FOUND IN THE BASEMENT
 CC MEMBRANES (MAJOR COMPONENT).
 CC -1- ALTERNATIVE PRODUCTS: THE TWO ISOFORMS A AND B, WHICH DIFFER
 CC IN THEIR N-TERMINAL ARE DERIVED BY ALTERNATIVE SPLICING OF THE
 CC SAME GENE. THE SEQUENCE SHOWN HERE IS THAT OF THE LARGER ISOFORM
 CC B.
 CC -1- TISSUE SPECIFICITY: BASAL MEMBRANE OF THE UPPER ALIMENTARY TRACT
 CC AND URINARY AND NASAL EPITHELIA, SALIVARY GLANDS AND TEETH (BOTH
 CC VARIANTS). ISOFORM A IS PREDOMINANTLY EXPRESSED IN SKIN, HAIR
 CC FOLLICLES AND DEVELOPING NEURONS OF THE TRIGEMINAL GANGLION.
 CC ISOFORM B WAS FOUND IN BRONCHI, ALVEOLI, STOMACH, INTESTINAL
 CC CRYPTS, WHISKER PADS, CNS, TELENCEPHALIC NEUROECTODERM, THALAMUS,
 CC RATHEE'S POUCH, AND PERIVENTRICULAR SUBPENDYMAL GERMINAL LAYER.
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -1- DOMAIN: DOMAINS IV AND G ARE GLOBULAR.
 CC -1- SIMILARITY: CONTAINS 6.5 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV (DOMAIN IV) IS NOT
 CC SIMILAR TO LAMININ DOMAIN IV.
 CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X84014; CA55837.1; -
 DR EMBL: X84013; CA55836.1; -
 DR EMBL: L20478; AAA68091.1; -
 DR HSSP: P02468; 1TLE.
 DR MSD: MGI:99909; LAMA3.
 DR INTERPRO: IPR000034; -
 DR INTERPRO: IPR000561; -
 DR INTERPRO: IPR001791; -
 DR INTERPRO: IPR002049; -
 DR PFAM: PF00052; laminin_B; 1.
 DR PFAM: PF00053; laminin_EGF; 4.
 DR PFAM: PF00054; laminin_G; 3.
 DR PROSITE: PS00022; EGF_1; 4.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; 4.
 DR Glycoprotein: Basement membrane: Extracellular matrix: Coiled coil;
 KM Laminin EGF-like domain; Cell adhesion; Repeat; Signal;
 KM Alternative splicing.
 FT SIGNAL 1
 FT NON_TER 28
 FT CHAIN 29 2569
 FT DOMAIN 29 498
 FT DOMAIN 499 700
 FT DOMAIN 700 700
 FT DOMAIN 546 700
 FT DOMAIN 546 589
 FT DOMAIN 590 639
 FT DOMAIN 640 690
 FT DOMAIN 691 700
 FT DOMAIN 701 889
 FT DOMAIN 890 1057
 FT DOMAIN 890 922
 FT DOMAIN 923 969
 FT DOMAIN 970 1022
 FT DOMAIN 1023 1057
 FT DOMAIN 1038 1648
 FT DOMAIN 1649 2569
 FT DOMAIN 1649 1825
 FT DOMAIN 1826 1994
 FT DOMAIN 1995 2209
 FT DOMAIN 2210 2385
 FT DOMAIN 2386 2569

FT DOMAIN 1090 1219 COILED COIL (POTENTIAL).
 FT DOMAIN 1251 1296 COILED COIL (POTENTIAL).
 FT DOMAIN 1327 1404 COILED COIL (POTENTIAL).
 FT DOMAIN 1450 1477 COILED COIL (POTENTIAL).
 FT DOMAIN 1557 1622 COILED COIL (POTENTIAL).
 FT SITE 1513 1515 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 546 553 BY SIMILARITY.
 FT DISULFID 548 560 BY SIMILARITY.
 FT DISULFID 562 571 BY SIMILARITY.
 FT DISULFID 574 587 BY SIMILARITY.
 FT DISULFID 590 605 BY SIMILARITY.
 FT DISULFID 592 612 BY SIMILARITY.
 FT DISULFID 614 623 BY SIMILARITY.
 FT DISULFID 626 637 BY SIMILARITY.
 FT DISULFID 640 652 BY SIMILARITY.
 FT DISULFID 642 659 BY SIMILARITY.
 FT DISULFID 661 670 BY SIMILARITY.
 FT DISULFID 673 688 BY SIMILARITY.
 FT DISULFID 923 932 BY SIMILARITY.
 FT DISULFID 925 939 BY SIMILARITY.
 FT DISULFID 942 951 BY SIMILARITY.
 FT DISULFID 954 967 BY SIMILARITY.
 FT DISULFID 970 982 BY SIMILARITY.
 FT DISULFID 972 991 BY SIMILARITY.
 FT DISULFID 993 1002 BY SIMILARITY.
 FT DISULFID 1005 1020 BY SIMILARITY.
 FT DISULFID 1058 1058 INTERCHAIN (PROBABLE).
 FT DISULFID 1061 1061 INTERCHAIN (PROBABLE).
 FT CARBOHYD 591 591 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 912 912 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1398 1398 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1500 1500 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1571 1571 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1600 1600 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1737 1737 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1819 1819 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1819 1819 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1986 1986 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2333 2333 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2509 2509 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 1 842 MISSING (IN ISOFORM A).
 FT VARSPLIC 843 901 MIVPLRIERHIGRIHTEFORALITGCVLEASDPGSGPR
 FT AHVENCACPDITGDSG -> MIPAYRMSKSTGWLITG
 FT AALGOCIGYSEQDQVAFIDRPPSONHILQASVMELRPS (IN
 FT ISOFORM A).
 SQ SEQUENCE 2569 AA; 282159 MW; 40D4306BER340DBC CRC64;
 Query Match 14.28; Score 79; DB 1; Length 2569;
 Best Local Similarity 29.48; Pred. No. 4.9;
 Matches 25; Conservative 7; Mismatches 35; Indels 18; Gaps 4;
 QY 21 GICVPCOINRYG---SVTNAGDLATLTQCTGCP-----TGTALDQSVTVDFR-- 67
 DB 937 GICINCQHNTRAGEHCERCQAGHYGNALHIGSCRCVCPHTNSFATGCAVDGAVACKPG 996
 QY 68 -SAAQCVKCKENFYNN---GGSPO 87
 DB 997 YTGQCERCACRGYGNPCKFGGSCQ 1021
 RESULT 2
 LMA5_MOUSE STANDARD; PRT; 3635 AA.
 ID LMA5_MOUSE
 AC 061001;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE LAMININ ALPHA-5 CHAIN (FRAGMENT).
 GN LAMA5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Mus.
 RN [1]


```
FT DISULFID 1486 1496 BY SIMILARITY.
FT DISULFID 1499 1511 BY SIMILARITY.
FT DISULFID 1501 1518 BY SIMILARITY.
FT DISULFID 1520 1529 BY SIMILARITY.
FT DISULFID 1532 1547 BY SIMILARITY.
FT DISULFID 1782 1791 BY SIMILARITY.
FT DISULFID 1784 1798 BY SIMILARITY.
FT DISULFID 1801 1810 BY SIMILARITY.
FT DISULFID 1813 1829 BY SIMILARITY.
FT DISULFID 1832 1847 BY SIMILARITY.
FT DISULFID 1834 1856 BY SIMILARITY.
FT DISULFID 1858 1867 BY SIMILARITY.
FT DISULFID 1870 1885 BY SIMILARITY.
FT DISULFID 1888 1903 BY SIMILARITY.
FT DISULFID 1890 1910 BY SIMILARITY.
FT DISULFID 1913 1922 BY SIMILARITY.
FT DISULFID 1925 1939 BY SIMILARITY.
FT DISULFID 1989 2000 BY SIMILARITY.
FT DISULFID 1991 2007 BY SIMILARITY.
FT DISULFID 2009 2018 BY SIMILARITY.
FT DISULFID 2021 2033 BY SIMILARITY.
FT DISULFID 2036 2043 BY SIMILARITY.
FT DISULFID 2038 2050 BY SIMILARITY.
FT DISULFID 2052 2061 BY SIMILARITY.
FT DISULFID 2064 2083 BY SIMILARITY.
FT DISULFID 2086 2086 INTERCHAIN (PROBABLE).
FT DISULFID 2089 2089 INTERCHAIN (PROBABLE).
FT CARBOHYD 17 17 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 65 65 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 402 402 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 822 822 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 843 843 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 881 881 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1252 1252 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1451 1451 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1938 1938 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2115 2115 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2128 2128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2282 2282 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2312 2312 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 14.0%; Score 78; DB 1; Length 3635;
Best Local Similarity 30.5%; Pred. No. 8.4;
Matches 25; Conservative 7; Mismatches 30; Indels 20; Gaps 4;

OY 21 GICVPCQINRGSVTN-----AGDLATLATQC-STQCP-----TGTALDDGVTD 63
    |||||
Db 1796 GICVGCQHNTEGDCRCRPGEVSSDPSNPASPCVSCPPLAVSNPNFADGCVLRNGRTQ 1855
    |||||

OY 64 VFDR---SAACVCKKPNFYNN 82
    |||||
Db 1856 CLCRPGYTAGASCERCARGPFRGN 1877
    |||||

RESULT 3
ASP_ANCCA STANDARD; PRT; 424 AA.
AC 016937;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ANCYLOSTOMA SECRETED PROTEIN PRECURSOR.
GN ASP.
OS Ancylostoma caninum (Dog hookworm).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
AN Ancylostomatoidae; Ancylostomatidae; Ancylostomatinae; Ancylostoma.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE-96215086; Pubmed-8636085.
RA Hawdon J.M., Jones B.F., Hoffman D.R., Hotez P.J.;
```

```
RT "Cloning and characterization of Ancylostoma-secreted protein. A
RT novel protein associated with the transition to parasitism by
RT infective hookworm larvae."
RL J. Biol. Chem. 271:6672-6678(1996).
CC -I- FUNCTION: ASSOCIATED WITH THE TRANSITION TO PARASITISM BY
CC INFECTIVE HOOKWORM LARVAE.
CC -I- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U26187; AAC47001.1; -.
DR HSSP; P04284; ICFE.
DR INTERPRO; IPR001283; -.
DR PFAM; PF00188; SCP. 1.
DR PRINTS; PR00837; VSTPLIKE.
DR PROSITE; PS01009; VSTP_LIKE.
DR PROSITE; PS01010; SCP_AGS_PRL_SC7_1; FALSE_NEG.
KW Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 424 ANCYLOSTOMA SECRETED PROTEIN.
FT SEQUENCE 424 AA; 45735 MW; 4BC8295F5D3035F9 CRC64;

Query Match 13.8%; Score 77; DB 1; Length 424;
Best Local Similarity 26.0%; Pred. No. 1.4;
Matches 26; Conservative 5; Mismatches 27; Indels 42; Gaps 4;

OY 8 NGNDPEANN-----AARGIC-----VPCQINRGSVTN-----AGDLATLATQC 47
    |||||
Db 142 NGGGLFAFSNNVYETTRKLCAYKVCGRKLAIVSGITVGNQPMWETGQACKTGADC 201
    |||||

OY 48 ST-----QCPGTALDDGVTDFV 65
    |||||
Db 202 STYKNSGCEDELCTKGPDPVPTNQCCPSNTGMTDSVTRTF 241
    |||||

RESULT 4
LMA_DROME STANDARD; PRT; 3712 AA.
AC 000174;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE LAMININ ALPHA CHAIN PRECURSOR.
GN LAMA OR LAMA.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93049203; Pubmed-1425586.
RA Kusche-Gulberg M., Garrison K., Mackrell A.J., Fessler L.I.,
RA Fessler J.H.;
RT "Laminin A chain: expression during Drosophila development and
RT genomic sequence."
RL EMBO J. 11:4519-4527(1992).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RX MEDLINE-94038678; Pubmed-8223265;
RA Henchcliffe C., Garcia-Alonso L., Tang J., Goodman C.S.;
RT "Genetic analysis of laminin A reveals diverse functions during
RT morphogenesis in Drosophila."
RL Development 118:325-337(1993).
[3]
```

	RP	SEQUENCE OF 1762-3712 FROM N.A.
	RX	MEDLINE=92078147; PubMed=1744083;
	RA	Garrison K., Mackrell A.J., Jessier J.H.:
	RT	"Drosophila laminin A chain sequence, interspecies comparison, and domain structure of a major carboxyl portion.";
	JL	J. Biol. Chem. 266:22899-22904(1991).
	CC	-I FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR. LAMININ IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
	CC	-I FUNCTION: DIVERSE FUNCTIONS DURING MORPHOGENESIS IN DROSOPHILA. COMPLETE LOSS-OF-FUNCTION MUTATIONS LEAD TO LATE EMBRYONIC LETHALITY. CERTAIN PARTIAL LOSS-OF-FUNCTION MUTATIONS GIVE RAISE TO ESCAPER ADULTS, WHICH HAVE ROUGH EGGS ASSOCIATED WITH CHANGES IN CELL FATE AND PATTERN, MISSHAPEN LEGS AND DEFECTS IN WING STRUCTURE.
	CC	-I SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
	CC	-I SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT MEMBRANES (MAJOR COMPONENT).
	CC	-I TISSUE SPECIFICITY: NEWLY FORMED MESODERM AND LATER PROMINENTLY EXPRESSED IN HEMOCYTES, WHICH ALSO SYNTHESIZE COLLAGEN IV.
	CC	-I DEVELOPMENTAL STAGE: DURING MORPHOGENESIS, MOSTLY IN EMBRIO DEVELOPMENT AT 10-12 HOURS.
	CC	-I DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
	CC	-I DOMAIN: DOMAINS VI, IV AND G ARE GLOBOULAR.
	CC	-I SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
	CC	-I SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.
	CC	-I SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV (DOMAIN IV) IT'S NOT SIMILAR TO LAMININ V.)
	CC	-I SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
	CC	-----
	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outpost - the European Bioinformatics institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/cor send an email to license@isb-sib.ch). -----
	DR	EMBL; M96388; AAA28662.1; -
	DR	EMBL; L07288; AAC37178.1; -
	DR	EMBL; M75882; AAC28661.1; -
	DR	HSSP; P02468; LTLE.
	DR	FLYBASE; FBgn00002526; Lama.
	DR	INTERPRO; IPR000034; -
	DR	INTERPRO; IPR000561; -
	DR	INTERPRO; IPR001791; -
	DR	INTERPRO; IPR001886; -
	DR	INTERPRO; IPR002049; -
	DR	Pfam; PF00052; laminin_B_1.
	DR	Pfam; PF00053; laminin_Egf_20.
	DR	Pfam; PF00054; laminin_G_5.
	DR	Pfam; PF00055; laminin_Nterm_1.
	DR	PRINTS; PRO0011; EGGFLAMIN.
	DR	PROSITE; PS00022; EGF_1; 17.
	DR	PROSITE; PS0186; EGF_2; 5.
	DR	PROSITE; PS01246; LAMININ_TYPB_EGF_19.
	KW	Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil; Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
	FT	SIGNAL 1..22 POTENTIAL.
	FT	CHAIN 23 .. 3712 LAMININ ALPHA CHAIN.
	FT	DOMAIN 25 .. 272 LAMININ N-TERMINAL (DOMAIN VI).
	FT	DOMAIN 273 .. 815 10..5 X LAMININ EGF-LIKE REPEATS (DOMAIN VII).
	FT	DOMAIN 273 .. 332 LAMININ EGF-LIKE 1.
	FT	DOMAIN 333 .. 402 LAMININ EGF-LIKE 2.
	FT	DOMAIN 403 .. 447 LAMININ EGF-LIKE 3.
	FT	DOMAIN 448 .. 494 LAMININ EGF-LIKE 4.
	FT	DOMAIN 495 .. 540 LAMININ EGF-LIKE 5.

FT	DOMAIN	541	586	LAMININ EGF-LIKE 6.
FT	DOMAIN	587	631	LAMININ EGF-LIKE 7.
FT	DOMAIN	632	676	LAMININ EGF-LIKE 8.
FT	DOMAIN	737	731	LAMININ EGF-LIKE 9.
FT	DOMAIN	772	784	LAMININ EGF-LIKE 10.
FT	DOMAIN	785	815	LAMININ EGF-LIKE 11 (INCOMPLETE).
FT	DOMAIN	816	1374	DOMAIN IV'.
FT	DOMAIN	1375	1574	4.5 X LAMININ EGF-LIKE REPEATS (DOMAIN III B).
FT	DOMAIN	1375	1420	LAMININ EGF-LIKE 12.
FT	DOMAIN	1421	1465	LAMININ EGF-LIKE 13.
FT	DOMAIN	1466	1513	LAMININ EGF-LIKE 14.
FT	DOMAIN	1514	1564	LAMININ EGF-LIKE 15.
FT	DOMAIN	1565	1574	LAMININ EGF-LIKE 16 (N-TERMINAL).
FT	DOMAIN	1575	1775	LAMININ DOMAIN IV (DOMAIN IV).
FT	DOMAIN	1776	2111	6.5 X LAMININ EGF-LIKE REPEATS (DOMAIN III A).
FT	DOMAIN	1776	1808	LAMININ EGF-LIKE 16 (C-TERMINAL).
FT	DOMAIN	1809	1858	LAMININ EGF-LIKE 17.
FT	DOMAIN	1859	1916	LAMININ EGF-LIKE 18.
FT	DOMAIN	1917	1969	LAMININ EGF-LIKE 19.
FT	DOMAIN	1970	2016	LAMININ EGF-LIKE 20.
FT	DOMAIN	2017	2063	LAMININ EGF-LIKE 21.
FT	DOMAIN	2064	2111	LAMININ EGF-LIKE 22.
FT	DOMAIN	2112	2697	DOMAIN II AND 1.
FT	DOMAIN	2698	3712	5 X LAMININ G-LIKE REPEATS (DOMAIN G).
FT	DOMAIN	2698	2862	LAMININ G-LIKE 1.
FT	DOMAIN	2863	3048	LAMININ G-LIKE 2.
FT	DOMAIN	3049	3223	LAMININ G-LIKE 3.
FT	DOMAIN	3270	3296	POLY-PHR.
FT	DOMAIN	3334	3528	LAMININ G-LIKE 4.
FT	DOMAIN	3529	3712	LAMININ G-LIKE 5.
FT	DOMAIN	2178	2249	COILED COIL (POTENTIAL).
FT	DOMAIN	2301	2321	COILED COIL (POTENTIAL).
FT	DOMAIN	2376	2450	COILED COIL (POTENTIAL).
FT	DOMAIN	2541	2676	COILED COIL (POTENTIAL).
FT	DISULFID	273	282	BY SIMILARITY.
FT	DISULFID	275	296	BY SIMILARITY.
FT	DISULFID	298	307	BY SIMILARITY.
FT	DISULFID	310	330	BY SIMILARITY.
FT	DISULFID	333	342	BY SIMILARITY.
FT	DISULFID	335	367	BY SIMILARITY.
FT	DISULFID	370	379	BY SIMILARITY.
FT	DISULFID	382	400	BY SIMILARITY.
FT	DISULFID	403	414	BY SIMILARITY.
FT	DISULFID	405	421	BY SIMILARITY.
FT	DISULFID	423	432	BY SIMILARITY.
FT	DISULFID	435	445	BY SIMILARITY.
FT	DISULFID	448	460	BY SIMILARITY.
FT	DISULFID	450	468	BY SIMILARITY.
FT	DISULFID	470	479	BY SIMILARITY.
FT	DISULFID	482	492	BY SIMILARITY.
FT	DISULFID	495	507	BY SIMILARITY.
FT	DISULFID	497	514	BY SIMILARITY.
FT	DISULFID	516	525	BY SIMILARITY.
FT	DISULFID	528	538	BY SIMILARITY.
FT	DISULFID	541	553	BY SIMILARITY.
FT	DISULFID	543	560	BY SIMILARITY.
FT	DISULFID	562	571	BY SIMILARITY.
FT	DISULFID	574	584	BY SIMILARITY.
FT	DISULFID	587	599	BY SIMILARITY.
FT	DISULFID	589	605	BY SIMILARITY.
FT	DISULFID	607	616	BY SIMILARITY.
FT	DISULFID	619	629	BY SIMILARITY.
FT	DISULFID	632	644	BY SIMILARITY.
FT	DISULFID	634	650	BY SIMILARITY.
FT	DISULFID	652	661	BY SIMILARITY.
FT	DISULFID	664	674	BY SIMILARITY.
FT	DISULFID	677	691	BY SIMILARITY.
FT	DISULFID	679	700	BY SIMILARITY.
FT	DISULFID	702	711	BY SIMILARITY.
FT	DISULFID	714	729	BY SIMILARITY.
FT	DISULFID	732	746	BY SIMILARITY.

```
FT DISULFID 734 753 BY SIMILARITY.
FT DISULFID 755 764 BY SIMILARITY.
FT DISULFID 767 782 BY SIMILARITY.
FT DISULFID 1375 1387 BY SIMILARITY.
FT DISULFID 1377 1394 BY SIMILARITY.
FT DISULFID 1396 1405 BY SIMILARITY.
FT DISULFID 1408 1418 BY SIMILARITY.
FT DISULFID 1421 1429 BY SIMILARITY.
FT DISULFID 1423 1436 BY SIMILARITY.
FT DISULFID 1438 1447 BY SIMILARITY.
FT DISULFID 1450 1463 BY SIMILARITY.
FT DISULFID 1466 1480 BY SIMILARITY.
FT DISULFID 1468 1487 BY SIMILARITY.
FT DISULFID 1489 1498 BY SIMILARITY.
FT DISULFID 1501 1511 BY SIMILARITY.
FT DISULFID 1514 1526 BY SIMILARITY.
FT DISULFID 1516 1533 BY SIMILARITY.
FT DISULFID 1535 1544 BY SIMILARITY.
FT DISULFID 1547 1562 BY SIMILARITY.
FT DISULFID 1859 1874 BY SIMILARITY.
FT DISULFID 1861 1885 BY SIMILARITY.
FT DISULFID 1887 1896 BY SIMILARITY.
FT DISULFID 1899 1914 BY SIMILARITY.
FT DISULFID 1917 1931 BY SIMILARITY.
FT DISULFID 1919 1938 BY SIMILARITY.
FT DISULFID 1941 1950 BY SIMILARITY.
```

Query Match 13.88; Score 77; DB 1; Length 3712;

Best Local Similarity 23.9%; Pred. No. 11; Matches 21; Conservative 17; Mismatches 20; Indels 30; Gaps 5;

```
OY 23 CVPQINRVGSGTNAGDLATLATIQCST-----QC-----PGTALD---DGV-TD 63
DB 492 CKACECKKISITNDCKVTGECCKLTNFGGDCNCRCKHGFNPTCSYCDNDQGTSE 551
OY 64 VFDSAAQCV-----KCKENFY 80
DB 552 ICKKSGQCTICREGGFGPRCDCLPGFY 579
```

```
RESULT 5
TSA4_GIALA STANDARD; PRT; 713 AA.
ID TSA4_GIALA
AC P21849;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE MAJOR SURFACE-LABELLED TROPHOZOITE ANTIGEN PRECURSOR.
GN TSA 417.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiae; Giardia.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-ATCC 30957 / WB:
RX MEDLINE=90280395; PubMed=2352929;
RX Gillin F.D., Hagblom P., Harwood J., Alez S.B., Reiner D.S.,
RA McCallery M., So M., Guiney D.G.;
RT Isolation and expression of the gene for a major surface protein of
RT Giardia lamblia."
RT Proc. Natl. Acad. Sci. U.S.A. 87:4463-4467(1990).
RN [2]
RP SEQUENCE OF 480-620 FROM N.A.
RX STRAIN-AD-1;
RX MEDLINE=93314970; PubMed=8325510;
RX Ey P.L., Mayrhofer G.;
RT "Two genes encoding homologous 70-kDa surface proteins are present
RT within individual trophozoites of the binucleate protozoan parasite
RT Giardia intestinalis."
RT Gene 129:257-262(1993).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- DOMAIN: CONTAINS 29 REPEATS OF THE CXXC MOTIF.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
```

Query Match 13.58; Score 75.5; DB 1; Length 713;

Best Local Similarity 27.5%; Pred. No. 3.2; Matches 28; Conservative 10; Mismatches 35; Indels 29; Gaps 6;

```
OY 3 AOGENGNGPFRANNAAMAGI--CVPQINRVGSGTNAGDLATLATIQC----- 47
DB 135 ACGDTGVTYIAAGNTYKGLADCAECSAPDATAGAEACKVAT-CTKCGVSRYLKDNVCVD 193
OY 48 STQCTPTF-----ALDDVTDFPDSAAQCVKCKENFYNG 84
DB 194 KAQCNSGSGTNKFAVVD-----SENGKVCVSCDNL--NGG 227
```

```
RESULT 6
LMG2_HUMAN STANDARD; PRT; 1193 AA.
ID LMG2_HUMAN
AC 013753; 013752; 014941; 002536; Q02537;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE LAMININ GAMMA-2 CHAIN PRECURSOR.
GN LAMC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=PLACENTA;
RX MEDLINE=96230326; PubMed=8786121;
RX Airenne T., Haekana H., Sainio K., Kallunki T., Kallunki P.,
RA Sariola H., Tryggvason K.;
RT "Structure of the human laminin gamma 2 chain gene (LAMC2):
RT alternative splicing with different tissue distribution of two
RT transcripts."
RT Genomics 32:54-64(1996).
RN [2]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 1090-1114.
RX TISSUE=EPIDERMIS, AND KERATINOCYTES;
RX MEDLINE=94139694; PubMed=8306988;
RX Vailly J., Verrando P., Champilaud M.F., Gerecke D., Wagnan D.W.,
RA Baudouin C., Aberdam D., Burgeson R., Bauer E., Ortonne J.P.;
RT "The 100-kDa chain of nectin/Kallinin is a laminin B2 chain variant."
RT Eur. J. Biochem. 219:209-218(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX TISSUE=FIBROSARCOMA;
RX MEDLINE=93016279; PubMed=1383240;
RX Kallunki P., Sainio K., Eddy R., Byers M., Kallunki T.,
RA Sariola H., Beck K., Hirvonen H., Tryggvason K.;
```



```

FT FT CARBOND 1033 1033 N-LINKED (GLCNAC...) (POTENTIAL).
FT FT VASAPPLIC 1110 1193 DOP1SYDEEGVILVLEOKTSRAKTQINISOLRPMWSPLEBRAR
OORGHLLHLETSIDGILADVKNLENTIRNLPPEGCTYTOALE
QO -> GM (IN SMALL ISOFORM).
FT FT CONFLICT 12 12 FQ -> L (IN REF. 2).
FT FT CONFLICT 473 473 M -> I (IN REF. 3).
FT FT CONFLICT 521 521 N -> S (IN REF. 3).
FT FT CONFLICT 857 857 R -> P (IN REF. 3).
FT FT CONFLICT 883 883 T -> S (IN REF. 2 AND 3).
SQ SEQUENCE 1193 AA; 130989 MW; 10BFC39E416FAECE CRC64;

Query Match 13.4%; Score 75; DB 1; Length 1193;
Best Local Similarity 26.7%; Pred. No. 5.9;
Matches 35; Conservative 11; Mismatches 25; Indels 60; Gaps 9.

QY 20 RGICPCQINRGSGS-----VTNAG--DIATLAT 45
D 78 RDRCLPCNCKNSGSLSARCNDNGRSCRPVYGARGCDRLCPHFHMLTDAGCTQDGRLLDS 137
QY 46 QCSITCPTGTALDDGVTVFEDRSAAQCYCKCP-----NFFYN--GSSPGGEADP 92
D 138 KCDCD-PAGIA--GPCD-----AGRCV-CRPATVGERCDRCRSGYVNLGDNPEB--C 184
QY 93 VQVFAAGAAAA 103
D 185 TQCFYGHSSAS 195

RESULT 7
LMAL_MOUSE STANDARD; PRT: 3084 AA.
ID LMAL_MOUSE
AC P19137;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE LAMININ ALPHA-1 CHAIN PRECURSOR (LAMININ A CHAIN).
GN LAMAI OR LAMA-1 OR LAMA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
LN 11
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RA Sasaki M., Kleitman H.K., Huber H., Deutzmann R., Yamada Y.;
RT "Laminin, a multidomain protein. The A chain has a unique globular
RT domain and homology with the basement membrane proteoglycan and the
RT laminin B chains."
RL J. Biol. Chem. 263:16536-16544(1988).
RN [2]
RN SEQUENCE OF 1-339 FROM N.A.
RP MEDLINE=88225080; PubMed=3267223;
RA Hartl L., Oberbauer I., Deutzmann R.;
RT "The N terminus of laminin A chain is homologous to the B chains."
RL Eur. J. Biochem. 173:629-635(1988).
RN [3]
RN SEQUENCE OF 2538-3084 FROM N.A., AND PARTIAL SEQUENCE.
RP MEDLINE=89030693; PubMed=3181157;
RA Deutzmann R., Huber J., Schmets R.A., Oberbauer I., Hartl L.;
RT "Structural study of long arm fragments of laminin. Evidence for the
RT repetitive C-terminal sequences in the A-chain, not present in the B-
RT chains."
RL Eur. J. Biochem. 177:35-45(1988).
RN [4]
RN FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
RN IS THOUGHT TO MEDIANE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
RN CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
RN WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
RN -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
RN DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
RN TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
RN COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
RN THE ALPHA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ) AND
RN LAMININ-3 (S-LAMININ).

```

[illegible]

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL: L34155; NAA59483.1; -
 DR EMBL: X85107; CAA59428.1; -
 DR EMBL: X85108; CAA59429.1; -
 DR HSSP: P02468; TITLE.
 DR MIM: 600805; -
 DR MIM: 226700; -
 DR INTERPRO: IPR000561; -
 DR INTERPRO: IPR001791; -
 DR INTERPRO: IPR002049; -
 DR PIRAM: PF00053; laminin_EGF; 2.
 DR PIRAM: PF00054; laminin_G; 2.
 DR PROSITE: PS00022; EGF_1; 1.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; 2.
 DR Glycoprotein: Basement membrane; Extracellular matrix; Coiled coil;
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal;
 KW Alternative splicing.
 KM SIGNAL.
 FT CHAIN 1 20
 FT DOMAIN 21 1713 POTENTIAL.
 FT DOMAIN 46 201 LAMININ ALPHA-3 CHAIN.
 FT DOMAIN 67 185 DOMAIN III A.
 FT DOMAIN 113 113 2.5 X LAMININ EGF-LIKE REPEATS.
 FT DOMAIN 114 166 LAMININ EGF-LIKE 1.
 FT DOMAIN 167 185 LAMININ EGF-LIKE 2.
 FT DOMAIN 202 793 LAMININ EGF-LIKE 3 (INCOMPLETE).
 FT DOMAIN 794 1713 DOMAIN II AND I (HEPAT REPEATS).
 FT DOMAIN 971 1139 5 X LAMININ G-LIKE REPEATS (DOMAIN G).
 FT DOMAIN 971 1139 LAMININ G-LIKE 1.
 FT DOMAIN 1140 1353 LAMININ G-LIKE 2.
 FT DOMAIN 1354 1529 LAMININ G-LIKE 3.
 FT DOMAIN 1530 1713 LAMININ G-LIKE 4.
 FT DOMAIN 231 327 LAMININ G-LIKE 5.
 FT DOMAIN 396 548 COILED COIL (POTENTIAL).
 FT DOMAIN 594 621 COILED COIL (POTENTIAL).
 FT DOMAIN 702 765 COILED COIL (POTENTIAL).
 FT DOMAIN 1686 1713 COILED COIL (POTENTIAL).
 FT DISULFID 67 76 BY SIMILARITY.
 FT DISULFID 69 83 BY SIMILARITY.
 FT DISULFID 86 95 BY SIMILARITY.
 FT DISULFID 98 111 BY SIMILARITY.
 FT DISULFID 114 126 BY SIMILARITY.
 FT DISULFID 116 135 BY SIMILARITY.
 FT DISULFID 137 146 BY SIMILARITY.
 FT DISULFID 149 164 BY SIMILARITY.
 FT DISULFID 202 202 INTERCHAIN (PROBABLE).
 FT DISULFID 205 205 INTERCHAIN (PROBABLE).
 FT SITE 658 660 CELL ATTACHMENT SITE (POTENTIAL).
 FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 745 745 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 882 882 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 964 964 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1108 1108 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1131 1131 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1325 1325 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1477 1477 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1667 1667 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VASPLIC 1 45 MGWITFGALGGLGYSOOQRYPIQPGOSLOASVE
 FT FRPS -> KVSSTIGITLYQKSSALDGDMLKRPVOL
 FT TGOHMSITVEETNPDPDRDLHGKRVHVEGFRHASSRAPY
 FT SREELATVLSRLADRIQGLYFTETORLTLEVGLEASPT
 FT GSGRIALVEICACPPAYAGDSC (IN ISOFORM B).
 FT W -> R (IN REF. 2).
 FT ATG -> GMC (IN REF. 2).
 FT M -> K (IN REF. 2).
 FT CONFLICT 5 5
 FT CONFLICT 123 125
 FT CONFLICT 481 481

FT CONFLICT 754 754 R -> L (IN REF. 2).
 FT CONFLICT 969 969 E -> Q (IN REF. 2).
 FT CONFLICT 1052 1052 D -> A (IN REF. 2).
 FT CONFLICT 1184 1184 G -> A (IN REF. 2).
 SQ SEQUENCE 1713 AA; 189304 MW; 45EA9BE1017B60D3 CRC64;

Query Match 12.7%; Score 71; DB 1; Length 1713;
 Best Local Similarity 30.0%; Pred. No. 20;
 Matches: 24; Conservative 6; Mismatches 22; Indels 28; Gaps 6;

QY 17 NAARGIC--VPCQINR---VGSVTNAGDLATLATCSTQCPTGYALDDGVYDPRSAQ 71
 DB 105 NAVHSCRCACPCPHNNSFATGCVVNGDV-----RCS--CKAG-----YTGTO 145
 QY 72 CYCKKPNFYNN-----GSGSQ 87
 DB 146 CERCAGYFGNPKRFGSQ 165

RESULT 10
 LMGI_MOUSE
 ID LMGI_MOUSE STANDARD: PRT: 1607 AA.
 AC P02468.
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE LAMININ GAMMA-1 CHAIN PRECURSOR (LAMININ B2 CHAIN).
 GN LAMC1 OR LAMC-1 OR LAMB-2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88059118; PubMed=3680290;
 RA Sasaki M., Yamada Y.;
 RT "The laminin B2 chain has a multidomain structure homologous to the
 RT B1 chain";
 RL J. Biol. Chem. 262:17111-17117(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8900737; PubMed=3167041;
 RA Durkin M.E., Bartos B.B., Liu S.-H., Phillips S.L., Chung A.E.;
 RT "Primary structure of the mouse laminin B2 chain and comparison with
 RT laminin B1";
 RL Biochemistry 27:5198-5204(1988).
 RN [3]
 RP SEQUENCE OF 1-239 FROM N.A.
 RX MEDLINE=88228071; PubMed=2836421;
 RA Ogawa K., Burdello P.D., Sasaki M., Yamada Y.;
 RT "The laminin B2 chain promoter contains unique repeat sequences and
 RT is active in transient transfection";
 RL J. Biol. Chem. 263:8384-8389(1988).
 RN [4]
 RP SEQUENCE OF 1391-1607 FROM N.A.
 RX MEDLINE=85051302; PubMed=6209134;
 RA Barlow D.P., Green N.M., Kurkinen M., Hogan B.L.M.;
 RT "Sequencing of laminin B chain cDNAs reveals C-terminal regions of
 RT coiled-coil alpha-helix";
 RL EMBO J. 3:2355-2362(1984).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 771-932.
 RX MEDLINE=96196434; PubMed=8648630;
 RA Stetefeld J., Mayer U., Timpl R., Huber R.;
 RT "Crystal structure of three consecutive laminin-type epidermal growth
 RT factor-like (LE) modules of laminin gamma1 chain harboring the
 RT nidogen binding site";
 RL J. Mol. Biol. 257:644-657(1996).
 RP [6]
 RP STRUCTURE BY NMR OF 824-881.
 RX MEDLINE=96196435; PubMed=8648631;
 RA Baumgartner R., Czisch M., Mayer U., Poeschl E., Huber R.,
 RA Timpl R., Holak T.A.;

"Structure of the nidogen binding LE module of the laminin gamma1 chain in solution." J. Mol. Biol. 257:658-668(1996).

-1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.

-1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END. THE GAMMA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ), LAMININ-2 (MEROSIN), LAMININ-3 (S-LAMININ), LAMININ-4 (S-MEROSIN), LAMININ-6 (K-LAMININ), AND LAMININ-7 (KS-LAMININ).

-1- SUBCELLULAR LOCATION: EXTRACELLULAR.

-1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR COMPONENT).

-1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.

-1- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.

-1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).

-1- SIMILARITY: CONTAINS 11 LAMININ EGF-LIKE DOMAINS.

-1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@sib-sib.ch).

CC -----

DR EMBL: X05211; CAA28838.1; -

DR EMBL: J03484; AAA39405.1; -

DR EMBL: J02930; AAA39408.1; -

DR EMBL: J03749; AAA39409.1; -

DR PIR: A28469; MMSB2.

DR PDB: 1KLO; 20-AUG-97.

DR MGD: MGI:99914; LAMC1.

DR INTERPRO: IPR000034; -

DR INTERPRO: IPR000561; -

DR INTERPRO: IPR001866; -

DR INTERPRO: IPR002049; -

DR PFAM: PF00052; laminin_B; 1.

DR PFAM: PF00053; laminin_EGF; 10.

DR PFAM: PF00055; laminin_Nterm; 1.

DR PRINTS: PR00011; EGF-LAMININ.

DR PROSITE: PS00022; EGF_1; 8.

DR PROSITE: PS01186; EGF_2; 2.

DR PROSITE: PS01248; LAMININ_TYPE_EGF; 10.

KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil; Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.

FT SIGNAL 1 33

FT CHAIN 34 1607

FT CHAIN 34 283

FT DOMAIN 284 502

FT 284 339

FT DOMAIN 340 395

FT DOMAIN 396 442

FT DOMAIN 443 492

FT DOMAIN 493 502

FT DOMAIN 503 687

FT DOMAIN 688 1028

FT 688 721

FT DOMAIN 722 770

FT DOMAIN 771 825

FT DOMAIN 826 881

FT DOMAIN 882 932

FT DOMAIN 933 980

FT DOMAIN 981 1028

LAMININ EGF-LIKE 1.

LAMININ EGF-LIKE 2.

LAMININ EGF-LIKE 3.

LAMININ EGF-LIKE 4.

LAMININ EGF-LIKE 5. (N-TERMINAL).

LAMININ DOMAIN IV.

6.5 X LAMININ EGF-LIKE REPEATS (DOMAIN III).

LAMININ EGF-LIKE 5 (C-TERMINAL).

LAMININ EGF-LIKE 6.

LAMININ EGF-LIKE 7.

LAMININ EGF-LIKE 8 (NIDOGEN-BINDING).

LAMININ EGF-LIKE 9.

LAMININ EGF-LIKE 10.

LAMININ EGF-LIKE 11.

FT	DOMAIN	1029	1607	DOMAIN II AND I.
FT	DOMAIN	1034	1594	COILED COIL (POTENTIAL).
FT	DISULFID	340	349	BY SIMILARITY.
FT	DISULFID	342	365	BY SIMILARITY.
FT	DISULFID	368	377	BY SIMILARITY.
FT	DISULFID	380	393	BY SIMILARITY.
FT	DISULFID	396	408	BY SIMILARITY.
FT	DISULFID	398	414	BY SIMILARITY.
FT	DISULFID	416	425	BY SIMILARITY.
FT	DISULFID	428	440	BY SIMILARITY.
FT	DISULFID	443	454	BY SIMILARITY.
FT	DISULFID	445	461	BY SIMILARITY.
FT	DISULFID	463	472	BY SIMILARITY.
FT	DISULFID	475	490	BY SIMILARITY.
FT	DISULFID	722	731	BY SIMILARITY.
FT	DISULFID	724	738	BY SIMILARITY.
FT	DISULFID	740	749	BY SIMILARITY.
FT	DISULFID	752	768	BY SIMILARITY.
FT	DISULFID	771	779	BY SIMILARITY.
FT	DISULFID	773	790	BY SIMILARITY.
FT	DISULFID	793	802	BY SIMILARITY.
FT	DISULFID	805	823	BY SIMILARITY.
FT	DISULFID	826	840	BY SIMILARITY.
FT	DISULFID	828	847	BY SIMILARITY.
FT	DISULFID	850	859	BY SIMILARITY.
FT	DISULFID	862	879	BY SIMILARITY.
FT	DISULFID	882	896	BY SIMILARITY.
FT	DISULFID	884	903	BY SIMILARITY.
FT	DISULFID	905	914	BY SIMILARITY.
FT	DISULFID	917	930	BY SIMILARITY.
FT	DISULFID	933	945	BY SIMILARITY.
FT	DISULFID	935	952	BY SIMILARITY.
FT	DISULFID	954	963	BY SIMILARITY.
FT	DISULFID	966	978	BY SIMILARITY.
FT	DISULFID	981	993	BY SIMILARITY.
FT	DISULFID	983	999	BY SIMILARITY.
FT	DISULFID	1001	1010	BY SIMILARITY.
FT	DISULFID	1013	1026	BY SIMILARITY.
FT	DISULFID	1029	1029	INTERCHAIN (PROBABLE).
FT	DISULFID	1032	1032	INTERCHAIN (PROBABLE).
FT	DISULFID	1598	1598	INTERCHAIN (WITH CHAIN BETA-1).
FT	CARBOHYD	58	58	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	132	132	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	574	574	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	648	648	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1020	1020	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1105	1105	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1159	1159	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1173	1173	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1203	1203	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1221	1221	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1239	1239	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1378	1378	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1393	1393	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1437	1437	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	216	216	G -> A (IN REF. 3).
FT	CONFLICT	260	260	E -> D (IN REF. 2).
FT	CONFLICT	337	337	S -> C (IN REF. 2).
FT	CONFLICT	447	448	LR -> PS (IN REF. 2).
FT	CONFLICT	544	544	D -> Y (IN REF. 2).
FT	CONFLICT	662	662	T -> S (IN REF. 2).
FT	CONFLICT	886	886	MISSING (IN REF. 2).
FT	CONFLICT	1158	1158	MISSING (IN REF. 2).
FT	CONFLICT	1434	1434	V -> A (IN REF. 2).
FT	CONFLICT	1475	1475	R -> K (IN REF. 4).
FT	CONFLICT	1576	1576	D -> N (IN REF. 4).
SO	SEQUENCE	1607	177297	MM; 81B7B08E4869E242 CRC64;

Query Match 12.6%; Score 70.5; DB 1; Length 1607;
Best Local Similarity 22.9%; Pred. No. 21;
Matches 25; Conservative 9; Mismatches 26; Indels 49; Gaps 5;

```

OY 23 CVPQINRVGVTNAGDLATLATGCTST-----QCPGTALD----- 59
DB 393 CSPCHSCSPVGS-----LSTGCDYGRSCRCRPGVMGDKDRCPGFHSLTEAGCRPC 443
OY 60 -----GYVDVFPBSAOCV-----KCPNFF-YNGSPGCEAP 91
DB 444 SCDLRGSTDCEWETGRCVCKNDVEGFCNCRCKPGPFMLESSNPGCTP 492

RESULT 11
LEM2_BOVIN
ID LEM2_BOVIN STANDARD: PRT: 485 AA.
AC P96107.
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 01-FEB-1996 (Rel. 33, Last annotation update)
E-ELECTIN PRECURSOR (ENDOTHELIAL LEUKOCYTE ADHESION MOLECULE 1)
DE (ELAM-1) (LEUKOCYTE-ENDOTHELIAL CELL ADHESION MOLECULE 2) (LECAM2)
DE (CD62E).
GN SELE.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-ADRENAL GLAND;
RX MEDLINE=93382537; PubMed=7690465;
RA Nguyen M., Strubel N.A., Bischoff J.;
RT "A role for sialyl Lewis-X/A glycoconjugates in capillary
morphogenesis.";
RL Nature 365:267-269(1993).
CC -1- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND
MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY
ELAM-1 IS SIALYL-LEWIS X (ALPHA1->3)FUCOSYLATED DERIVATIVES OF
POLYLACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF
GLYCOPOLYIDS).
CC -1- FUNCTION: INVOLVED IN CAPILLARY MORPHOGENESIS BY BINDING TO A
BOVINE CAPILLARY ENDOTHELIAL (BCE) CELL SIALYL LEWIS-X AND/OR
SIALYL LEWIS A-CONTAINING LIGAND (PROBABLY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: TO OTHER SELECTINS/LECAMs.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS; BOVINE E-LECTIN LACKS
THE HUMAN SUSHI-4 AND 5 EQUIVALENTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L12039; AAA02991.1; -
DR HSSP; P16581; IESL.
DR INTERPRO: IPR000436; -
DR INTERPRO: IPR000561; -
DR INTERPRO: IPR001304; -
DR INTERPRO: IPR002396; -
DR PFAM; PF00008; EGF_1.
DR PFAM; PF00059; lectin_c; 1.
DR PFAM; PF00084; sush1; 4.
DR PRINTS; PR00343; SELECTIN.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00615; C-TYPE_LLECTIN_1; 1.
DR PROSITE; PS50041; C-TYPE_LLECTIN_2; 1.
KW Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW Selectin; Signal; Sush1; Repeat.
FT SIGNAL 1 22 BY SIMILARITY.
FT CHAIN 23 485 E-SELECTIN.

```

```

FT DOMAIN 23 430 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 431 453 POTENTIAL.
FT DOMAIN 454 485 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 39 139 C-TYPE LECTIN (SHORT FORM).
FT DOMAIN 140 176 EGF-LIKE.
FT DOMAIN 180 422 4 X SUSHI (SCR) REPEATS.
FT REPEAT 180 238 SUSHI 1.
FT REPEAT 241 300 SUSHI 2.
FT REPEAT 303 363 SUSHI 3.
FT REPEAT 366 422 SUSHI 4.
FT DISULFD 41 139. BY SIMILARITY.
FT DISULFD 112 131. BY SIMILARITY.
FT DISULFD 144 155. BY SIMILARITY.
FT DISULFD 149 164. BY SIMILARITY.
FT DISULFD 166 175. BY SIMILARITY.
FT DISULFD 181 224. BY SIMILARITY.
FT DISULFD 210 237. BY SIMILARITY.
FT DISULFD 242 286. BY SIMILARITY.
FT DISULFD 272 299. BY SIMILARITY.
FT DISULFD 304 349. BY SIMILARITY.
FT DISULFD 335 362. BY SIMILARITY.
FT DISULFD 367 408. BY SIMILARITY.
FT DISULFD 394 421. BY SIMILARITY.
FT CARBOHYD 61 61. N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 79 79. N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 88 88. N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 161 161. N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 203 263. N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 265 265. N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 312 312. N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 316 316. N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 379 379. N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 401 401. N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 485 AA; 53200 MW; AE931C9B521E3904 CRC64;

Query Match 12.5%; Score 70; DB 1; Length 485;
Best Local Similarity 27.2%; Pred. No. 7.8;
Matches 25; Conservative 12; Mismatches 35; Indels 20; Gaps 4;

OY 24 VP-CQINRVGVTNAGDL-----ATLATGCTSTGCPGTALDGYVDVFPBSAOCVYC 75
DB 359 VPCEVWVRCSTRLDVGSKLNMNCGSEPVYGTCTCPACPRMTLNGSV-----LTC 408
OY 76 KENPFYNGSPGCEAPGYQ--VFAAGAAAGV 105
DB 409 GATGHWGMLPTCEAPFYVQTPLAVGLSTAGV 440

RESULT 12
LML1_CAEL
ID LML1_CAEL STANDARD: PRT: 1557 AA.
AC Q18823;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE LAMININ-LIKE PROTEIN C54D1.5 PRECURSOR.
GN C54D1.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Mlnx P.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -1- SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U46673; AAC48152.1; -
 DR HSSP: P02468; 1LE.
 DR WORMPEP: C54D1.5; CE06981.
 DR INTERPRO: IPR000534; -
 DR INTERPRO: IPR000561; -
 DR INTERPRO: IPR00186; -
 DR INTERPRO: IPR002049; -
 DR PFAM: PF00052; laminin_B; 1.
 DR PFAM: PF00053; laminin_EGF; 10.
 DR PFAM: PF00055; laminin_Nterm; 1.
 DR PRINTS: PR00011; EGF_LAMININ.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS00022; EGF_1; 8.
 DR PROSITE: PS01248; LAMININ_EGF-like domain; signal; Repeat.
 DR PROSITE: PS01248; LAMININ_EGF-like domain; signal; Repeat.
 KW Hypothetical protein; Laminin EGF-like domain; signal; Repeat.
 FT SIGNAL 1 1557
 FT CHAIN ? 271
 FT DOMAIN ? 271
 FT DOMAIN 272 331
 FT DOMAIN 332 387
 FT DOMAIN 388 434
 FT DOMAIN 435 487
 FT DOMAIN 488 497
 FT DOMAIN 498 688
 FT DOMAIN 687 732
 FT DOMAIN 737 769
 FT DOMAIN 770 819
 FT DOMAIN 820 874
 FT DOMAIN 875 930
 FT DOMAIN 931 978
 FT DOMAIN 979 1025
 FT DOMAIN 1025 1025
 FT DISULFID 272 281
 FT DISULFID 274 295
 FT DISULFID 297 306
 FT DISULFID 309 329
 FT DISULFID 332 341
 FT DISULFID 334 357
 FT DISULFID 360 369
 FT DISULFID 372 385
 FT DISULFID 388 400
 FT DISULFID 390 406
 FT DISULFID 408 417
 FT DISULFID 420 432
 FT DISULFID 435 449
 FT DISULFID 437 456
 FT DISULFID 458 467
 FT DISULFID 470 485
 FT DISULFID 470 778
 FT DISULFID 772 788
 FT DISULFID 791 800
 FT DISULFID 803 817
 FT DISULFID 820 834
 FT DISULFID 822 841
 FT DISULFID 844 853
 FT DISULFID 856 872
 FT DISULFID 875 894
 FT DISULFID 877 901
 FT DISULFID 903 912
 FT DISULFID 915 928
 FT DISULFID 931 943
 FT DISULFID 933 950
 FT DISULFID 952 961
 FT DISULFID 964 976
 FT DISULFID 979 991
 FT DISULFID 981 998
 FT DISULFID 1000 1009
 FT DISULFID 1012 1023
 FT CARBOHYD 101 101

FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 507 507 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 643 643 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 921 921 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1062 1062 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1168 1168 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1211 1211 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1244 1244 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1321 1321 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1437 1437 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1557 AA; 172723 MW; CA0B51F8D5E8D2F CRC64;

Query Match 12.4% Score 69; DB 1; Length 1557;
 Best Local Similarity 23.4% Pred. No. 29;
 Matches 29; Conservative 8; Mismatches 31; Indels 56; Gaps 6;

Qy 4 QGEANG--NQPFAANMAR--GICVPCQINRVSVTWAGDLATLATQCSF---QCP-- 52
 Db 361 QGNNGVHCCECIANHHMRPEENCVCAGCMEIS-----LSIQCDNEGKCCQCPG 411
 Qy 53 -GTALDGDGVTVDFRS-----AAQCYKCK 76
 Db 412 VTGRFCDCLDGFDFSTNGCKNGCETSGSLNQPDCSSSSGSCCKLNVGRQCKCK 471
 Qy 77 PNFY 80
 Db 472 PGIF 475

RESULT 13.
 TNRC_MOUSE
 ID TNRC_MOUSE STANDARD; PRT; 415 AA.
 AC P50284; 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE LYMPHOTOXIN-BETA RECEPTOR PRECURSOR.
 GN LYBR OR TNFR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-CVB; TISSUE=LUNG;
 RX MEDLINE=96072804; PubMed=7594541;
 RA Force W.R., Walter B.N., Hession C., Tizard R., Kozak C.A.,
 RA Browning J.L., Ware C.F.;
 RA "Mouse lymphotoxin-beta receptor. Molecular genetics, ligand binding,
 RA and expression";
 RL J. Immunol. 155:5280-5288(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96163885; PubMed=8586432;
 RA Nakamura T., Tashiro K., Nazarea M., Nakano T., Sasayama S.,
 RA Honjo T.;
 RA "The murine lymphotoxin-beta receptor cDNA: isolation by the signal
 RA sequence trap and chromosomal mapping";
 RL Genomics 30:312-319(1995).
 CC -1- FUNCTION: RECEPTOR FOR THE LYMPHOTOXIN-BETA. POSSIBLE FUNCTION IN
 CC IMMUNE DEVELOPMENT.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- SIMILARITY: CONTAINS A LA-NGR/TNFR-TYPE CYSTEINE-RICH REGION.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

Db 712 CLNCVHSLGSKVT-----SRKCVSVCPICGDTAARCRCHKCGCTCSSRAATQC 764

QY 73 VKCRPNFYNN 82

Db 765 LSCRGFYHH 774

RESULT 15

SL17_ENTHI STANDARD; PRT; 1285 AA.

AC P23502;

DT 01-NOV-1991 (Rel. 20, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 01-FEB-1994 (Rel. 28, Last annotation update)

DE 170 KDA SURFACE LECTIN PRECURSOR.

GN CEL-170/4.

OS Entamoeba histolytica.

OC Eukaryota; Entamoebidae; Entamoeba.

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-33.

RX MEDLINE=91156704; PubMed=2000392;

RA Tannich E., Ebert F., Horstmann R.D.;

RT "Primary structure of the 170-kda surface lectin of pathogenic

RL Entamoeba histolytica.";

RL Proc. Natl. Acad. Sci. U.S.A. 88:1849-1853(1991).

RN [2]

RP REVISIONS TO N-TERMINUS.

RA Tannich E., Nickel R., Horstmann R.D.;

RL Submitted (AUG-1991) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 16-30.

RX MEDLINE=89123408; PubMed=2536731;

RA Petri W.A. Jr., Chapman M.D., Snodgrass T., Mann B.J., Broman J.,

RA Raydin J.I.;

RT "Subunit structure of the galactose and N-acetyl-D-galactosamine-

RT inhibitable adherence lectin of Entamoeba histolytica.";

RL J. Biol. Chem. 264:3007-3012(1989).

CC -1- FUNCTION: MEDIATES ADHERENCE OF E. HISTOLYTICA TO COLONIC MUCINS,

CC AN ESSENTIAL STEP FOR PATHOGENIC TISSUE INVASION.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -1- PTM: PHOSPHORYLATED REGIONS MAY HAVE A ROLE IN SIGNAL

CC TRANSDUCTION.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M60498; AAA29106.1; -

DR EMBL; X61003; CAA43321.1; -

DR PIR; A39117; A39117.

DR PIR; S21942; S21942.

KW Signal; Lectin; Transmembrane; Repeat; Glycoprotein;

KW Phosphorylation.

FT SIGNAL 1 15

FT CHAIN 16 1285

FT DOMAIN 16 1218

FT TRANSMEM 1219 1247

FT DOMAIN 1248 1285

FT DOMAIN 378 1208

FT DOMAIN 378 653

FT REPEAT 378 407

FT REPEAT 408 438

FT REPEAT 439 464

FT REPEAT 465 497

FT REPEAT 498 525

FT REPEAT 526 554

FT REPEAT 555 584

FT REPEAT 585 620

FT REPEAT 621 653

FT 170 KDA SURFACE LECTIN.

FT EXTRACELLULAR (POTENTIAL).

FT POTENTIAL.

FT CYTOPLASMIC (POTENTIAL).

FT CYS-RICH.

FT 9 X APPROXIMATE TANDEM REPEATS.

FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 236 236 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 383 383 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 461 461 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 882 882 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1205 1205 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 16 16 D -> G (IN REF. 3).

SQ SEQUENCE 1285 AA; 144410 MW; 64BB2DFD2AD8ECBA CRC64;

Query Match 12.3%; Score 68.5; DB 1; Length 1285;

Best Local Similarity 28.4%; Pred. No. 27;

Matches 25; Conservative 10; Mismatches 26; Indels 27; Gaps 6;

QY 7 ANGNOFFAANNAARGICVP-----COINRYGSTNMGDLATLATGCTGCTGPTGA 56

Db 665 SNTKIEFAQDKSFTMCKPYSATCLNGCQVQAAGDVSNVG-----CG-YCSMGT- 714

QY 57 LDDGVT--DVFPDRAAOC---VKCRPN 78

Db 715 -DNVITTHDDCDSDKRSQCNGNKGKCPN 741

Search completed: March 6, 2001, 12:54:31

Job time: 388 sec

THIS PAGE BLANK (USPTO)

XX This invention describes novel isolated Streptococcus pneumoniae
 CC polynucleotides (see Z96173-Z96494) and their encoded proteins (see
 CC Y85792-Y86182). The DNA, vectors and host cells described in the method
 CC of the invention are useful for the recombinant expression of the
 CC polypeptides. The polypeptides are useful for treatment or prevention of
 CC disease, or diagnosis of disease related to expression or activity of
 CC such a polypeptide. They can also be used to screen for compounds which
 CC interact with and inhibit or activate such a polypeptide. The
 CC polypeptides (or DNA encoding them, via gene therapy) are also useful
 CC for inducing an immunological response in a mammal. The antagonists are
 CC useful to inhibit such bacterial polypeptides. The polypeptides are
 CC particularly useful to identify antimicrobial compounds and antibiotics.
 CC They are also useful to determine their role in pathogenesis of
 CC infection, dysfunction and disease.

XX
 SQ Sequence 200 AA:

Query Match 100.0%; Score 33; DB 19; Length 200;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEANGN 6
 111111
 Db 166 geangn 171

RESULT 2

W98367
 ID W98367 standard; Protein; 791 AA.

XX W98367;

DT 31-MAR-1999 (first entry)

DE H. pylori GHPO 359 protein.

KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
 KM peptic ulcer disease.

XX Helicobacter pylori.

PN W09843478-A1.

PD 08-OCT-1998.

PF 01-APR-1998; 98WO-US06371.

PR 29-JUL-1997; 97US-0902615.

PR 01-APR-1997; 97US-0833457.

PR 24-JUN-1997; 97US-0881227.

PA (HUMA-) HUMAN GENOME SCI INC.
 (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;

DR WPI: 1998-542293/46.

DR N-PSDB; X14086.

PT New isolated Helicobacter polynucleotides - used to develop products
 for the diagnosis, prevention and treatment of Helicobacter

PS Claim 8; Page 553-556; 2054pp; English.

XX This sequence represents a Helicobacter pylori GHPO protein of the
 CC invention. The polypeptides can be used for preventing or treating
 CC Helicobacter infections, and gastroduodenal diseases associated with
 CC these infections, including acute, chronic, and atrophic gastritis, and
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
 CC used for the production of antibodies. The products can also be used for

CC detection and diagnosis.

XX
 SQ Sequence 791 AA;

Query Match 87.9%; Score 29; DB 19; Length 791;
 Best Local Similarity 83.3%; Pred. No. 3.5e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEANGN 6
 111111
 Db 462 gkangn 467

RESULT 3

W81504
 ID W81504 standard; Protein; 1079 AA.

XX W81504;

DT 02-FEB-1999 (first entry)

DE Short form of TPR motif Y (TPRY) gene product.

KW Non-recombining region; human; Y chromosome; X homologue; testis; TPRY;
 KM infertility; sperm; gene alteration; inhibitor.

XX Homo sapiens.

OS

PN W09846747-A2.

PD 22-OCT-1998.

PF 10-APR-1998; 98WO-US07115.

PR 11-APR-1997; 97US-0041877.

PA (WHED) WHITEHEAD INST BIOMEDICAL RES.

PI Lahn BT, Page DC;

DR WPI: 1998-568729/48.

DR N-PSDB; V69634.

PT Novel genes in the non-combining region of Y chromosome - useful to
 PT diagnose if male infertility or reduced sperm count has a genetic
 PT basis

PS Claim 6; Fig 4A-B; 54pp; English.

XX This represents the amino acid sequence of the short form of the TPR
 CC motif Y (TPRY) gene product. The invention relates to genes occurring on
 CC the non-recombining region of the human Y chromosome. The sequences fall
 CC into two classes: (1) X-homologous DNA which are expressed in many
 CC organs, having functional X homologues and (2) testis-specific DNA
 CC sequences. Y chromosomal DNA from males with known conditions such as
 CC infertility and reduced sperm count can be assessed using the invention
 CC to determine whether the condition is associated with or caused by the
 CC occurrence of the gene or gene alteration. Candidate inhibitors of the
 CC enzymatic activity of the genes can be assessed using in vitro assays.

XX
 SQ Sequence 1079 AA;

Query Match 87.9%; Score 29; DB 19; Length 1079;
 Best Local Similarity 83.3%; Pred. No. 4.5e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEANGN 6
 111111
 Db 733 gkangn 738

RESULT 4
 ID W81505 standard; Protein; 1240 AA.
 AC W81505;
 XX
 XX
 DT 02-FEB-1999 (first entry)
 DE Medium form of TPR motif Y (TPRY) gene product.
 XX
 KW Non-recombining region; human; Y chromosome; X homologue; testis; TPRY;
 KM Infertility; sperm; gene alteration; inhibitor.
 XX
 OS Homo sapiens.
 PN W09846747-A2.
 PD 22-OCT-1998.
 PF 10-APR-1998; 98WO-US07115.
 PR 11-APR-1997; 97US-0041877.
 PS
 PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
 PI Lahn BT, Page DC;
 XX
 XX WPI; 1998-568729/48.
 DR N-PSDB; V69635.
 XX
 PT Novel genes in the non-combining region of Y chromosome - useful to
 PT diagnose if male infertility or reduced sperm count has a genetic
 PT basis
 PS
 PS Claim 6; Fig 4A-B; 54pp; English.
 CC This represents the amino acid sequence of the medium form of the TPR
 CC motif Y (TPRY) gene product. The invention relates to genes occurring on
 CC the non-recombining region of the human Y chromosome. The sequences fall
 CC into two classes: (1) X-homologous DNA which are expressed in many
 CC organs, having functional X homologues and (2) testis-specific DNA
 CC sequences. Y chromosomal DNA from males with known conditions such as
 CC infertility and reduced sperm count can be assessed using the invention
 CC to determine whether the condition is associated with or caused by the
 CC occurrence of the gene or gene alteration. Candidate inhibitors of the
 CC enzymatic activity of the genes can be assessed using in vitro assays.
 CC
 SO Sequence 1240 AA.

Query Match 87.9%; Score 29; DB 19; Length 1240;
 Best Local Similarity 83.3%; Pred. No. 5.7e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEANGN 6
 ID W81505 standard; Protein; 1347 AA.
 AC W81506;
 XX
 XX
 DT 02-FEB-1999 (first entry)
 DE Long form of TPR motif Y (TPRY) gene product.
 XX
 KW Non-recombining region; human; Y chromosome; X homologue; testis; TPRY;
 KM Infertility; sperm; gene alteration; inhibitor.
 XX
 OS Homo sapiens.
 PI

XX W09846747-A2.
 XX
 XX 22-OCT-1998.
 PD
 XX
 XX 10-APR-1998; 98WO-US07115.
 PF
 XX 11-APR-1997; 97US-0041877.
 PR
 XX
 PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
 PI Lahn BT, Page DC;
 XX
 XX WPI; 1998-568729/48.
 DR N-PSDB; V69636.
 XX
 PT Novel genes in the non-combining region of Y chromosome - useful to
 PT diagnose if male infertility or reduced sperm count has a genetic
 PT basis
 PS
 PS Claim 6; Fig 4A-B; 54pp; English.
 CC This represents the amino acid sequence of the long form of the TPR
 CC motif Y (TPRY) gene product. The invention relates to genes occurring on
 CC the non-recombining region of the human Y chromosome. The sequences fall
 CC into two classes: (1) X-homologous DNA which are expressed in many
 CC organs, having functional X homologues and (2) testis-specific DNA
 CC sequences. Y chromosomal DNA from males with known conditions such as
 CC infertility and reduced sperm count can be assessed using the invention
 CC to determine whether the condition is associated with or caused by the
 CC occurrence of the gene or gene alteration. Candidate inhibitors of the
 CC enzymatic activity of the genes can be assessed using in vitro assays.
 CC
 SO Sequence 1347 AA;

Query Match 87.9%; Score 29; DB 19; Length 1347;
 Best Local Similarity 83.3%; Pred. No. 6.2e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEANGN 6
 ID W82618 standard; Protein; 185 AA.
 AC W82618;
 XX
 XX
 DT 01-MAR-1999 (first entry)
 DE Ehrlichia sp. HGE-7 protein N-terminal.
 XX
 XX Antigen: human granulocytic ehrlichiosis; HGE; diagnosis; treatment;
 KM immunogenic; infection; vaccine; epitope; medication; rodent; bacteria;
 KM protective immunity; Lyme disease; babesiosis; detection; HGE-7.
 XX
 OS Ehrlichia sp.
 PN W09842740-A2.
 PD 01-OCT-1998.
 PF 23-MAR-1998; 98WO-US05695.
 PR 20-NOV-1997; 97US-0975762.
 PR 21-MAR-1997; 97US-0821324.
 PA (CORI-) CORIXA CORP.
 PI Houghton R, Iodes MJ, Reed SG;

XX DR WPI; 1998-609891/51.
 DR N-PSDB; V69406.
 XX
 PT Poly:peptide(s) comprising immunogenic portion of Ehrlichia antigen
 PT - and encoding DNA sequences, useful for e.g. diagnosis and
 PT treatment of Ehrlichia infection, especially human granulocytic
 PT ehrlichiosis
 XX
 PS Example 1; Page 38; 140pp; English.
 CC This sequence represents an immunogenic portion of Ehrlichia antigen,
 CC HGE-7. This polypeptide is useful in the treatment of Ehrlichia
 CC infection, and as a vaccine for the prevention of infection. Such
 CC vaccines comprise an immunogenic portion of an Ehrlichia antigen
 CC associated with human granulocytic ehrlichiosis (or a variant) and is
 CC thus especially useful in the treatment of human granulocytic
 CC ehrlichiosis (HGE). The polypeptides, antigenic epitopes or DNA
 CC molecules can be combined with a suitable carrier in pharmaceutical
 CC compositions. Such compositions and vaccines are useful to manufacture
 CC medicaments for inducing protective immunity against Ehrlichia infection
 CC in patients especially against HGE. HGE is caused by a rodent bacterium
 CC normally transmitted to humans by the same tick which transmits Lyme
 CC disease and babesiosis. Co-infection with these diseases is thus possible
 CC and the compositions of the invention may be used in methods to detect at
 CC least one of Ehrlichia infection, Lyme disease or B. microti infection in
 CC patients.
 CC
 XX
 SO Sequence 185 AA;

Query Match 84.8%; Score 28; DB 19; Length 185;
 Best Local Similarity 83.3%; Pred. No. 1.2e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GEANGN 6
 II III
 DB 92 geqnqn 97

RESULT 7
 Y78527
 ID Y78527 standard; Protein; 185 AA.
 XX
 AC Y78527;
 XX
 DT 05-MAY-2000 (first entry)
 XX
 DE Ehrlichia antigen 7 (HGE-7) protein sequence.
 XX
 KW Ehrlichia antigen; diagnose; infection; vaccine; Lyme disease; HGE;
 KM human granulocytic ehrlichiosis; Babesia microti infection; HGE-7.
 XX
 OS Ehrlichia sp.
 XX
 PN WO200000615-A2.
 XX
 PD 06-JAN-2000.
 XX
 PF 29-JUN-1999; 99WO-US14793.
 XX
 PR 29-JUN-1998; 98US-0106582.
 PR 23-SEP-1998; 98US-0159469.
 PR 20-APR-1999; 99US-0295028.
 XX
 PA (CORI-) CORIXA CORP.
 PI Reed SG, Lodes MJ, Houghton RL, McNeill PD;
 XX WPI; 2000-160675/14.
 DR N-PSDB; 289971.
 XX
 PT New compounds and methods for the diagnosis of Ehrlichia infection,

PT particularly Human granulocytic ehrlichiosis
 XX
 PS Claim 2; Page 54; 109pp; English.
 XX
 CC This sequence represents the amino acid sequence of an Ehrlichia antigen
 CC HGE-7. The invention relates to immunogenic portions of Ehrlichia
 CC antigens which are used in the production of new compounds and methods
 CC for the diagnosis of Ehrlichia infections. The antigen polypeptides,
 CC nucleotide sequences encoding the antigens, a fusion protein comprising
 CC at least one antigen and/or antigenic epitope, are useful for the
 CC detection and treatment of Ehrlichia infections. The proteins,
 CC polynucleotides, antigenic epitopes and fusion proteins can also be used
 CC to detect Lyme disease and Babesia microti infections. In particular,
 CC the polypeptides containing at least two antigenic epitopes can be used
 CC for the sero-diagnosis and treatment of human granulocytic ehrlichiosis
 CC (HGE). Compositions of the polypeptides, nucleotide sequences, and
 CC antigenic epitopes can be used for the manufacture of a medicament for
 CC inducing protective immunity in a patient. New vaccines can also be used
 CC for inducing protective immunity in a patient.
 CC
 XX
 SO Sequence 185 AA;

Query Match 84.8%; Score 28; DB 21; Length 185;
 Best Local Similarity 83.3%; Pred. No. 1.2e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GEANGN 6
 II III
 DB 92 geqnqn 97

RESULT 8
 W82637
 ID W82637 standard; Protein; 253 AA.
 XX
 AC W82637;
 XX
 DT 01-MAR-1999 (first entry)
 XX
 DE Ehrlichia sp. HGE-7 protein #1.
 XX
 KW Antigen; human granulocytic ehrlichiosis; HGE; diagnosis; treatment;
 KM immunogenic; infection; vaccine; epitope; medicament; rodent; bacteria;
 KM protective immunity; Lyme disease; babesiosis; detection; HGE-7.
 XX
 OS Ehrlichia sp.
 XX
 PN WO9842740-A2.
 XX
 PD 01-OCT-1998.
 XX
 PF 23-MAR-1998; 98WO-US05695.
 XX
 PR 20-NOV-1997; 97US-0975762.
 PR 21-MAR-1997; 97US-0821324.
 XX
 PA (CORI-) CORIXA CORP.
 PI Houghton R, Lodes MJ, Reed SG;
 XX WPI; 1998-609891/51.
 DR N-PSDB; V69422.
 XX
 PT Poly:peptide(s) comprising immunogenic portion of Ehrlichia antigen
 PT - and encoding DNA sequences, useful for e.g. diagnosis and
 PT treatment of Ehrlichia infection, especially human granulocytic
 PT ehrlichiosis
 XX
 PS Example 1; Page 94-95; 140pp; English.
 CC This sequence represents an immunogenic portion of Ehrlichia antigen,
 CC HGE-7. This polypeptide is useful in the treatment of Ehrlichia

CC infection, and as a vaccine for the prevention of infection. Such
CC vaccines comprise an immunogenic portion of an Ehrlichia antigen
CC associated with human granulocytic ehrlichiosis (or a variant) and is
CC thus especially useful in the treatment of human granulocytic
CC ehrlichiosis (HGE). The polypeptides, antigenic epitopes or DNA
CC molecules can be combined with a suitable carrier in pharmaceutical
CC compositions. Such compositions and vaccines are useful to manufacture
CC medicaments for inducing protective immunity against Ehrlichia infection
CC in patients especially against HGE. HGE is caused by a rodent bacterium
CC normally transmitted to humans by the same tick which transmits Lyme
CC disease and babesiosis. Co-infection with these diseases is thus possible
CC and the compositions of the invention may be used in methods to detect at
CC least one of Ehrlichia infection, Lyme disease or B. microti infection in
CC patients.

CC Sequence 253 AA;

Query Match 84.8%; Score 28; DB 19; Length 253;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GEANGN 6
|||
DB 92 gqgnqn 97

RESULT 9

ID Y78528 standard; Protein; 253 AA.

AC Y78528;

DT 05-MAY-2000 (first entry)

XX Ehrlichia antigen 7 (HGE-7) protein sequence #2.

XX Ehrlichia antigen; diagnose; infection; vaccine; Lyme disease; HGE;

KW human granulocytic ehrlichiosis; Babesia microti infection; HGE-7.

OS Ehrlichia sp.

PN W0200000615-A2.

PD 06-JAN-2000.

XX 29-JUN-1999; 99MO-US14793.

PR 29-JUN-1998; 98US-0106582.

PR 23-SEP-1998; 98US-0159469.

PR 20-APR-1999; 99US-0295028.

XX (CORI-) CORIXA CORP.

PI Reed SG, Lodes MJ, Houghton RL, McNeill PD;

DR WPI: 2000-160675/14.

XX N-PSDB; Z89987.

PT New compounds and methods for the diagnosis of Ehrlichia infection,

XX particularly Human granulocytic ehrlichiosis

PS Claim 2; Page 84; 109pp; English.

CC This sequence represents the amino acid sequence of an Ehrlichia antigen
CC HGE-7. The invention relates to immunogenic portions of Ehrlichia
CC antigens which are used in the production of new compounds and methods
CC for the diagnosis of Ehrlichia infections. The antigen polypeptides,
CC nucleotide sequences encoding the antigens, a fusion protein comprising
CC at least one antigen and/or antigenic epitope, are useful for the
CC detection and treatment of Ehrlichia infections. The proteins,
CC polynucleotides, antigenic epitopes and fusion proteins can also be used
CC to detect Lyme disease and Babesia microti infections. In particular,

CC the polypeptides containing at least two antigenic epitopes can be used
CC for the sero-diagnosis and treatment of human granulocytic ehrlichiosis
CC (HGE). Compositions of the polypeptides, nucleotide sequences, and
CC antigenic epitopes can be used for the manufacture of a medicament for
CC inducing protective immunity in a patient. New vaccines can also be used
CC for inducing protective immunity in a patient.

CC Sequence 253 AA;

Query Match 84.8%; Score 28; DB 21; Length 253;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GEANGN 6
|||
DB 92 gqgnqn 97

RESULT 10
ID Y38784 standard; Protein; 297 AA.

AC Y38784;

DT 08-OCT-1999 (first entry)

XX Neisseria gonorrhoeae antigenic protein encoded by ORF138.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.

OS Neisseria gonorrhoeae.

PN W09924578-A2.

PD 20-MAY-1999.

XX 09-OCT-1998; 98MO-IB01665.

PR 01-SEP-1998; 98GB-0019016.

PR 06-NOV-1997; 97GB-0023516.

PR 14-NOV-1997; 97GB-0024190.

PR 18-NOV-1997; 97GB-0024386.

PR 27-NOV-1997; 97GB-0025158.

PR 10-DEC-1997; 97GB-0026147.

PR 14-JAN-1998; 98GB-0000759.

XX (CHIR-) CHIRON SPA.

PI Grandi G, Masignani V, Pizza M, Rappelli R, Scarlato V;

DR WPI: 1999-327407/27.

XX N-PSDB; Z12219.

PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for

XX diagnosis, treatment and prevention of infection

PS Claim 4; Page 328; 524pp; English.

CC Amino acid sequences Y38499-Y38944 represent Neisseria meningitidis
CC and N. gonorrhoeae antigenic proteins. They are encoded by open
CC reading frames (ORFs) Z11972-Z12358. The antigenic proteins,
CC their fragments, their nucleic acids and antibodies are used for
CC diagnosis, prevention (as vaccines) or treatment of Neisseria
CC infections, such as meningitis, septicaemia and gonorrhea. Both
CC organisms are closely related. Fragments of the nucleic acids
CC are useful as hybridisation probes and antisense reagents.

Query Match 84.8%; Score 28; DB 20; Length 297;

Best Local Similarity 83.3%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GEANGN 6
|||
Db 259 gelngn 264

RESULT 11

W74947
ID W74947 standard; Protein; 297 AA.

XX AC W74947;

XX DT 21-MAR-2000 (first entry)

XX DE Neisseria gonorrhoeae ORF 505 protein sequence SEQ ID NO:1368.

XX KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
XX KM antibacterial; gene therapy.

XX OS Neisseria gonorrhoeae.

XX PN W09957280-A2.

XX PD 11-NOV-1999.

XX PF 30-APR-1999; 99MO-US09346.

XX PR 01-MAY-1998; 98US-0083758.

XX PR 31-JUL-1998; 98US-0094869.

XX PR 02-SEP-1998; 98US-0098994.

XX PR 02-SEP-1998; 98US-0099062.

XX PR 09-OCT-1998; 98US-0103749.

XX PR 09-OCT-1998; 98US-0103794.

XX PR 09-OCT-1998; 98US-0103796.

XX PR 25-FEB-1999; 99US-0121528.

XX PA (CHIR) CHIRON CORP.

XX PI (GENO-) INST GENOMIC RES.

XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
XX PI Petersen J, Pilza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX PI Tettelin H, Venter JC;

XX DR WPI; 2000-062150/05.

XX DR N-PSDB; 253709.

XX PT Novel Neisserial polypeptides predicted to be useful antigens for
XX PT vaccines and diagnostics

XX PS Claim 2; Page 744; 1453pp; English.

XX CC 253015 to 254536, 254577 to 254615, and W74253 to W75941 represent
XX CC novel Neisseria meningitis and N. gonorrhoeae polynucleotides and
XX CC polypeptides. 254537 to 254576 and 254616 to 25473 represent PCR
XX CC primers used in the exemplification of the present invention. The
XX CC polypeptides, the polynucleotides, antibodies and compositions of
XX CC the invention can be used as vaccines, as diagnostic reagents, and as
XX CC immunogenic compositions. The polypeptides can be used in the
XX CC manufacture of medicaments for treating or preventing infection due to
XX CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
XX CC presence of Neisseria bacteria, or to raise antibodies. They may also
XX CC have use as antibacterial agents. The polynucleotides of the invention
XX CC may also be used in gene therapy protocols.

XX SO Sequence 297 AA;

Query Match 84.8%; Score 28; DB 21; Length 297;
Best Local Similarity 83.3%; Pred. No. 2e+02;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GEANGN 6
|||
Db 259 gelngn 264

RESULT 12

W67442
ID W67442 standard; Protein; 568 AA.

XX AC W67442;

XX DT 02-MAR-1999 (first entry)

XX DE Bacillus species alpha-glucosidase.

XX KM Alpha-glucosidase; Bacillus species KS-108a; Km; dissociation constant;
XX KW specificity; stability; recombinant; transgenic organism.

XX OS Bacillus sp.

XX PN JP10286092-A.

XX PD 27-OCT-1998.

XX PF 15-APR-1997; 97JP-0111778.

XX PR 15-APR-1997; 97JP-0111778.

XX PA (KIKK) KIKKOMAN CORP.

XX DR WPI; 1999-017004/02.

XX DR N-PSDB; V34379.

XX PT New alpha-glucosidase gene - useful for recombinant production of
XX PT alpha glucosidase with, e.g. low Km value

XX PS Claim 1; Page 6-7; 8pp; Japanese.

XX CC This sequence represents a novel alpha-glucosidase from a Bacillus sp.
XX CC strain KS-108a. The novel alpha-glucosidase has a low Km value and
XX CC excellent substrate specificity and stability. The protein can be
XX CC produced recombinantly in a transgenic host organism.

XX SO Sequence 568 AA;

Query Match 84.8%; Score 28; DB 20; Length 568;
Best Local Similarity 83.3%; Pred. No. 3.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEANGN 6
|||
Db 254 geangns 259

RESULT 13

W55103
ID W55103 standard; Protein; 711 AA.

XX AC W55103;

XX DT 02-OCT-1998 (first entry)

XX DE Streptococcus pneumoniae SP0051 protein.

XX KM Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
XX KW detection; pneumonia; otitis media; meningitis.

XX OS Streptococcus pneumoniae.

XX FH Key Location/Qualifiers


```

FT Misc-difference 56
FT /label= "unknown"
FT /note= "encoded by NAG"
XX
XX
XX WO9818930-A2.
XX
XX PD 07-MAY-1998.
XX
XX PF 30-OCT-1997; 97WO-US19422.
XX
XX PR 31-OCT-1996; 96US-0029960.
XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Chol GH, Hromocky J A, Johnson LS, Kunsch CA;
XX
XX DR WPI: 1998-272224/24.
XX
XX DR N-PSDB: VZ7364.
XX
XX PT Nucleic acid encoding antigenic peptide(s) from Streptococcus
XX PT pneumoniae - or their epitope-containing fragments, useful in
XX PT protective or therapeutic vaccines, and for diagnosis
XX
XX PS Claim 11; Page 66; 118pp; English.
XX
XX CC The present sequence represents a protein from Streptococcus pneumoniae.
XX CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
XX CC can be useful in vaccines for inducing protective antibodies against
XX CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
XX CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
XX CC are used to detect Streptococcus infection (by usual hybridisation or
XX CC amplification methods), also for isolating Streptococcus genes or their
XX CC allelic variants. The protein can be used similarly to detect specific
XX CC antibodies in standard immunoassays, especially for diagnosing or
XX CC monitoring infections. Antibodies which bind the protein are used to
XX CC detect corresponding antigens, to purify the protein and for passive
XX CC immunisation (optionally coupled to a toxin). Vaccines are administered,
XX CC e.g. by injection, orally or through the skin, typically at 0.01-1000
XX CC (especially 10-300) mu g/ml per dose.
XX
XX SO Sequence 711 AA;

Query Match 84.8%; Score 28; DB 19; Length 711;
Best Local Similarity 83.3%; Pred. No. 5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEANGN 6
Db 526 geangd 531

RESULT 14
R13793
ID R13793 standard; Protein; 878 AA.
XX
XX AC R13793;
XX
XX DT 29-NOV-1991 (first entry)
XX
XX DE Ecdysone receptor.
XX
XX KW Insect steroid receptor; ECR; hormone; DHR23.
XX
XX OS Drosophila melanogaster.
XX
XX FH Key
XX FH Location/Qualifiers
XX FT 264..329
XX FT /note= "zinc-finger DNA-binding domain C"
XX FT Domain
XX FT 431..651
XX FT /note= "hormone-binding domain E"
XX
XX PN WO9113167-A.

```

```

XX
XX PD 05-SEP-1991.
XX
XX PF 15-FEB-1991; 91WO-US01189.
XX
XX PR 26-FEB-1990; 90US-0485749.
XX
XX PA (STRD ) IELAND STANFORD JR UNIV.
XX
XX PI Hogness DS, Koelle MR, Segraves WA;
XX
XX DR WPI: 1991-281480/38.
XX
XX DR N-PSDB: Q13574.
XX
XX PT DNA encoding insect steroid receptors - and ligands, for use as
XX PT benign inducing factors
XX
XX PS Claim 24; Page 103; 126pp; English.
XX
XX CC The amino acid sequence codes for the (20-OH) ecdysone receptor
XX CC protein which is part of the insect steroid receptor superfamily.
XX CC It can be used to screen for ligands specific for the insect
XX CC steroid receptors which can be used as highly specific and highly
XX CC active pesticides which are biodegradable. See also R13791-R13794.
XX
XX SQ Sequence 878 AA;

Query Match 84.8%; Score 28; DB 12; Length 878;
Best Local Similarity 83.3%; Pred. No. 6.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEANGN 6
Db 100 gnangn 105

RESULT 15
R32889
ID R32889 standard; Protein; 878 AA.
XX
XX AC R32889;
XX
XX DT 22-JUN-1993 (first entry)
XX
XX DE DHR23alpha protein.
XX
XX KW Ecdysteroid; receptor; DHR23alpha; Drosophila; steroid; receptor;
XX KW ligand-dependant; transcription factor; mammalian cell; hormone;
XX KW ecdysone response element; ecdysone.
XX
XX OS Drosophila melanogaster.
XX
XX FH Key
XX FH Location/Qualifiers
XX FT Binding-site
XX FT 281..330
XX FT /note= "Putative DNA-binding domain"
XX
XX PN WO9303162-A.
XX
XX PD 18-FEB-1993.
XX
XX PF 03-AUG-1992; 92WO-US06391.
XX
XX PR 08-AUG-1991; 91US-0742127.
XX
XX PA (GETH ) GENENTECH INC.
XX
XX PI Godowski PJ;
XX
XX DR WPI: 1993-076519/09.
XX
XX DR N-PSDB: Q37556.
XX
XX PT Ecdysteroid dependent regulation of genes in mammalian cells -

```

PT for induced expression of heterologous genes

PS Disclosure; Fig 1; 45pp; English.

CC This sentence represents an ecdysteroid receptor DHR23alpha. DHR23-
CC alpha is a Drosophila steroid receptor homologue which can function as
CC a ligand-dependant transcription factor in mammalian cells when induced
CC by specific ecdysteroids such as DHR23alpha. The activity of DHR23-
CC alpha is induced upon administration of certain ecdysteroids but not
CC any of the mammalian hormones tested. DHR23alpha has been reported to
CC regulate transcription of genes containing ecdysone response elements
CC in Drosophila tissue culture cells treated with 20-OH ecdysone.

SQ Sequence 878 AA;

Query Match	84.88;	Score 28;	DB 14;	Length 878;
Best Local Similarity	83.38;	Prod No 630403;		

Best Local Similarity 83.38; Pred. No. 6.3e+02;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GEANGN 6

Db 100 gnangn 105

Search completed: March 6, 2001, 12:49:32
Job time: 93 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OK protein - protein search, using sw model

Run on: March 6, 2001, 12:53:46 ; Search time 57.76 seconds
(without alignments)
1.865 Million cell updates/sec

Title: US-09-196-161D-3
Perfect score: 33
Sequence: 1 GEANGN 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/PCRTUS.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	30	90.9	200	2 US-08-682-517-20	Sequence 20, Appl
2	29	87.9	344	1 US-08-446-919A-4	Sequence 4, Appl
3	29	87.9	1079	3 US-09-058-489-22	Sequence 22, Appl
4	29	87.9	1240	3 US-09-058-489-23	Sequence 23, Appl
5	29	87.9	1347	3 US-09-058-489-24	Sequence 24, Appl
6	28	84.8	711	3 US-08-961-083-82	Sequence 82, Appl
7	27	81.8	10	3 US-09-101-167-2	Sequence 2, Appl
8	27	81.8	108	2 US-08-440-354-4	Sequence 4, Appl
9	27	81.8	108	2 US-08-463-087-4	Sequence 4, Appl
10	27	81.8	535	1 US-08-039-777-3	Sequence 3, Appl
11	27	81.8	555	1 US-08-611-361A-3	Sequence 3, Appl
12	27	81.8	572	2 US-08-453-848-9	Sequence 9, Appl
13	27	81.8	592	3 US-08-961-083-222	Sequence 222, App
14	27	81.8	659	1 US-08-258-639A-4	Sequence 4, Appl
15	27	81.8	659	2 US-08-900-951-4	Sequence 4, Appl
16	27	81.8	659	4 PCT-US95-07391A-4	Sequence 4, Appl
17	27	81.8	663	4 US-08-776-265-5	Sequence 4, Appl
18	27	81.8	749	1 US-08-369-796-8	Sequence 8, Appl
19	27	81.8	749	2 US-08-852-091-8	Sequence 8, Appl
20	27	81.8	749	2 US-08-820-754-8	Sequence 8, Appl
21	27	81.8	749	3 US-08-956-652-8	Sequence 8, Appl
22	27	81.8	749	3 US-08-956-659-8	Sequence 8, Appl
23	27	81.8	749	3 US-09-012-710-7	Sequence 8, Appl
24	27	81.8	749	3 US-08-948-547-8	Sequence 8, Appl
25	27	81.8	749	4 PCT-US95-17025-8	Sequence 8, Appl
26	26	78.8	198	2 US-08-682-517-19	Sequence 19, Appl
27	26	78.8	1222	2 US-08-682-517-15	Sequence 15, Appl
28	26	78.8	1252	2 US-08-682-517-9	Sequence 9, Appl

29	25	75.8	97	3 US-08-894-699-26	Sequence 26, Appl
30	25	75.8	159	3 US-09-027-449-53	Sequence 53, Appl
31	25	75.8	159	3 US-08-804-444A-53	Sequence 53, Appl
32	25	75.8	159	3 US-09-026-985-53	Sequence 53, Appl
33	25	75.8	211	1 US-08-276-882-34	Sequence 34, Appl
34	25	75.8	211	1 US-08-133-011-16	Sequence 16, Appl
35	25	75.8	211	1 US-08-322-730A-16	Sequence 16, Appl
36	25	75.8	211	1 US-08-387-874-16	Sequence 16, Appl
37	25	75.8	211	1 US-08-899-575-34	Sequence 34, Appl
38	25	75.8	211	1 US-08-899-575-34	Sequence 34, Appl
39	25	75.8	211	2 US-08-383-619-16	Sequence 16, Appl
40	25	75.8	211	4 PCT-US93-08364-16	Sequence 16, Appl
41	25	75.8	211	4 PCT-US95-08743-34	Sequence 34, Appl
42	25	75.8	232	2 US-08-829-876-103	Sequence 103, Appl
43	25	75.8	233	2 US-08-829-876-101	Sequence 101, Appl
44	25	75.8	233	2 US-08-829-876-105	Sequence 105, Appl
45	25	75.8	233	4 PCT-US94-05669A-15	Sequence 15, Appl

ALIGNMENTS

```

RESULT 1
US-08-682-517-20
: Sequence 20, Application US/08682517
: Patent No. 5874267
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: Expression of surface layer proteins
: NUMBER OF SEQUENCES: 25
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/682,517
: FILING DATE:
: CLASSIFICATION:
: INFORMATION FOR SEQ ID NO: 20:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 200 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE:
US-08-682-517-20

Query Match          90.9%; Score 30; DB 2; Length 200;
Best Local Similarity 83.3%; Pred. No. 51;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEANGN 6
I:||||
Db 58 GDANGN 63

RESULT 2
US-08-446-919A-4
: Sequence 4, Application US/08446919A
: Patent No. 5736389
: GENERAL INFORMATION:
: APPLICANT: Kiozler, Kenneth W.
: TITLE OF INVENTION: Vogelstein, Bert
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Banner & Allegretti, Ltd.
: STREET: 1001 G Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.

```

ZIP: 20001-4597
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,919A
FILING DATE:
CLASSIFICATION: 544
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 01107.49255
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202.508.9100
TELEFAX: 202.508.9299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 344 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharomyces cerevisiae
IMMEDIATE SOURCE:
CLONE: Yero16p
US-08-446-919A-4

Query Match 87.9%; Score 29; DB 1; Length 344;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GEANGN 6
11111
Db 310 GKANGN 315

RESULT 3
US-09-058-489-22
Sequence 22, Application US/09058489
Patent No. 6103886
GENERAL INFORMATION:
APPLICANT: Whitehead Institute for Biomedical Research
APPLICANT: Lahn, Bruce
TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
FILE REFERENCE: WHI97-08pA
CURRENT APPLICATION NUMBER: US/09/058,489
CURRENT FILING DATE: 1998-04-10
EARLIER APPLICATION NUMBER: 60/041,877
EARLIER FILING DATE: 1997-04-11
NUMBER OF SEQ ID NOS: 91
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 22
LENGTH: 1079
TYPE: PRT
ORGANISM: Human
US-09-058-489-22

Query Match 87.9%; Score 29; DB 3; Length 1079;
Best Local Similarity 83.3%; Pred. No. 4.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEANGN 6
11111
Db 733 GKANGN 738

RESULT 4
US-09-058-489-23
Sequence 23, Application US/09058489
Patent No. 6103886
GENERAL INFORMATION:
APPLICANT: Whitehead Institute for Biomedical Research
APPLICANT: Lahn, Bruce
TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
FILE REFERENCE: WHI97-08pA
CURRENT APPLICATION NUMBER: US/09/058,489
CURRENT FILING DATE: 1998-04-10
EARLIER APPLICATION NUMBER: 60/041,877
EARLIER FILING DATE: 1997-04-11
NUMBER OF SEQ ID NOS: 91
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 23
LENGTH: 1240
TYPE: PRT
ORGANISM: Human
US-09-058-489-23

Query Match 87.9%; Score 29; DB 3; Length 1240;
Best Local Similarity 83.3%; Pred. No. 5.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEANGN 6
11111
Db 733 GKANGN 738

RESULT 5
US-09-058-489-24
Sequence 24, Application US/09058489
Patent No. 6103886
GENERAL INFORMATION:
APPLICANT: Whitehead Institute for Biomedical Research
APPLICANT: Lahn, Bruce
TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
FILE REFERENCE: WHI97-08pA
CURRENT APPLICATION NUMBER: US/09/058,489
CURRENT FILING DATE: 1998-04-10
EARLIER APPLICATION NUMBER: 60/041,877
EARLIER FILING DATE: 1997-04-11
NUMBER OF SEQ ID NOS: 91
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 24
LENGTH: 1347
TYPE: PRT
ORGANISM: Human
US-09-058-489-24

Query Match 87.9%; Score 29; DB 3; Length 1347;
Best Local Similarity 83.3%; Pred. No. 5.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEANGN 6
11111
Db 733 GKANGN 738

RESULT 6
US-08-961-083-82
Sequence 82, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961.083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 82:
SEQUENCE CHARACTERISTICS:
LENGTH: 711 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-82

Query Match 84.8%; Score 28; DB 3; Length 711;
Best Local Similarity 83.3%; Pred. No. 4.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEANGN 6
Db 526 GEANGD 531

RESULT 7
US-09-101-167-2
Sequence 2, Application US/09101167B
Patent No. 6127139
GENERAL INFORMATION:
APPLICANT: TE KOPPELE, Johannes M
APPLICANT: BEERMAN, Bob
TITLE OF INVENTION: METHOD FOR ASSAYING PROTEOLYTIC ENZYMES USING
TITLE OF INVENTION: FLUORESCENCE-QUENCHED SUBSTRATES
FILE REFERENCE: TE KOPPELE et al. 09/101.167
CURRENT APPLICATION NUMBER: US/09/101.167B
CURRENT FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: EP 96200017.0
EARLIER FILING DATE: 1996-01-04
EARLIER APPLICATION NUMBER: PCT/NL97/00002
EARLIER FILING DATE: 1997-01-02
NUMBER OF SEQ ID NOS: 43
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2
LENGTH: 10
TYPE: PRT
ORGANISM: unknown
FEATURE:
OTHER INFORMATION: Description of Unknown Organism: UNKNOWN
US-09-101-167-2

Query Match 81.8%; Score 27; DB 3; Length 108;
Best Local Similarity 83.3%; Pred. No. 9.3;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GEANGN 6
Db 48 GEANGN 53

RESULT 8
US-08-440-354-4
Sequence 4, Application US/08440354
Patent No. 5907034
GENERAL INFORMATION:
APPLICANT: Boblet, Klaus
APPLICANT: Hermentin, Peter
APPLICANT: Seemann, Gerhard
TITLE OF INVENTION: A Monoclonal Antibody Against Complexed
TITLE OF INVENTION: and No. 5907034Complexed Complexing Agents for Removing Hea
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440.354
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/013.166
FILING DATE: 02-FEB-1993
APPLICATION NUMBER: US/07/664.789
FILING DATE: 05-MAR-1991
APPLICATION NUMBER: DE 4007079.4
FILING DATE: 07-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Lavin Jr., Lawrence M.
REGISTRATION NUMBER: 30,768
REFERENCE/DOCKET NUMBER: HOAG-1043
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 108 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-440-354-4

Query Match 81.8%; Score 27; DB 2; Length 108;
Best Local Similarity 83.3%; Pred. No. 1e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GEANGN 6
Db 48 GPANGN 53

RESULT 9
US-08-463-087-4

Sequence 4, Application US/08463087
Patent No. 5908790
GENERAL INFORMATION:
APPLICANT: Bobiet, Klaus
APPLICANT: Hermentlin, Peter
APPLICANT: Seemann, Gerhard
TITLE OF INVENTION: A Monoclonal Antibody Against Complexed
TITLE OF INVENTION: and No. 5908790Complexed Complexing Agents for Removing Heavy
TITLE OF INVENTION: Metals from Aqueous Solutions and for Analysis
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,087
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/664,789
FILING DATE: 05-MAR-1991
APPLICATION NUMBER: DE 4007079.4
FILING DATE: 07-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Layin Jr., Lawrence M.
REGISTRATION NUMBER: 30,768
REFERENCE/DOCKET NUMBER: HOAG-1043
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 108 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-463-087-4

Query Match 81.8%; Score 27; DB 2; Length 108;
Best Local Similarity 83.3%; Pred. No. 1e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GEANGN 6
| | | | |
Db 48 GPANGN 53

RESULT 10
US-08-039-777-3
Sequence 3, Application US/08039777
Patent No. 5530046
GENERAL INFORMATION:
APPLICANT: Suzuki, Yuzuru
APPLICANT: Takii, Yukio
APPLICANT: Yamamoto, Kazumi
APPLICANT: Nishiya, Yoshiaki
APPLICANT: Sogabe, Atushi
APPLICANT: Sogabe, Yukihiko
APPLICANT: Eml, Shigenori
TITLE OF INVENTION: PROTEIN HAVING -GLUCOSIDASE ACTIVITY, DNA HAVING
TITLE OF INVENTION: GENETIC INFORMATION THEREOF, AND PRODUCTION OF -GLUCOSIDASE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:

ADDRESSEE: Leydig, Voit & Mayer
STREET: Two Prudential Plaza, Suite 4900
CITY: Chicago
STATE: Illinois
COUNTRY: U.S.A.
ZIP: 60601-6780
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/039,777
FILING DATE: 19930322
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Green, Robert F.
REGISTRATION NUMBER: 27555
REFERENCE/DOCKET NUMBER: 56250
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)616-5600
TELEFAX: (312)616-5700
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 555 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-039-777-3

Query Match 81.8%; Score 27; DB 1; Length 555;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEANG 5
| | | | |
Db 255 GEANG 259

RESULT 11
US-08-611-361A-3
Sequence 3, Application US/08611361A
Patent No. 5795766
GENERAL INFORMATION:
APPLICANT: Suzuki, Yuzuru
APPLICANT: Takii, Yukio
APPLICANT: Yamamoto, Kazumi
APPLICANT: Nishiya, Yoshiaki
APPLICANT: Sogabe, Atushi
APPLICANT: Sogabe, Yukihiko
APPLICANT: Eml, Shigenori
TITLE OF INVENTION: PROTEIN HAVING -GLUCOSIDASE ACTIVITY, DNA
TITLE OF INVENTION: HAVING GENETIC INFORMATION THEREOF, AND PRODUCTION OF
TITLE OF INVENTION: -GLUCOSIDASE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Leydig, Voit & Mayer
STREET: Two Prudential Plaza, Suite 4900
CITY: Chicago
STATE: Illinois
COUNTRY: U.S.A.
ZIP: 60601-6780
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/611,361A
FILING DATE: 06-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 92101658
FILING DATE: 27-MAR-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 92117538
FILING DATE: 11-MAY-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Green, Robert F.
REGISTRATION NUMBER: 27555
REFERENCE/DOCKET NUMBER: 56250
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)616-5600
TELEFAX: (312)616-5700
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 555 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-611-361A-3

Query Match 81.8%; Score 27; DB 1; Length 555;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GEANG 5
|||||
Db 255 GEANG 259

RESULT 12
US-08-453-848-9
Sequence 9, Application US/08453848
Patent No. 5838368
GENERAL INFORMATION:
APPLICANT: Smith, Gale Eugene
APPLICANT: Volvovitz, Franklin
APPLICANT: Wilkinson, Bethanie Eident
APPLICANT: Voznesensky, Andrei I.
APPLICANT: Hackett, Craig Stanley
TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA
TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: GA
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/453,848
FILING DATE: 30-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/120,607
FILING DATE: 13-SEPT-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: MGS101CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404)-873-8794
TELEFAX: (404)-873-8795

INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 572 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Influenza virus
INDIVIDUAL ISOLATE: A/Texas/36/91 rHA
FEATURE:
NAME/KEY: ACNPV 61k protein signal sequence
LOCATION: 1 to 18
FEATURE:
NAME/KEY: mature rHA
LOCATION: 19 to 554
US-08-453-848-9

Query Match 81.8%; Score 27; DB 2; Length 572;
Best Local Similarity 100.0%; Pred. No. 5.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 EANGN 6
|||||
Db 268 EANGN 272

RESULT 13
US-08-961-083-222
Sequence 222, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 222:
SEQUENCE CHARACTERISTICS:
LENGTH: 592 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-222

Query Match 81.8%; Score 27; DB 3; Length 592;
Best Local Similarity 100.0%; Pred. No. 5.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 EANGN 6
DB 556 EANGN 560

RESULT 14

US-08-258-639A-4
; Sequence 4, Application US/08258639A
; Patent No. 5681733
; GENERAL INFORMATION:
; APPLICANT: Su, Hongsheng
; APPLICANT: Blain, Francoise
; APPLICANT: Bennett, Clark
; APPLICANT: Gu, Kangfu
; APPLICANT: Zimmermann, Joseph
; APPLICANT: Musil, Roy
; TITLE OF INVENTION: Nucleic Acid Sequences And Expression
; TITLE OF INVENTION: Systems For Heparinase II And Heparinase III Derived From
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Hale and Dorf
; STREET: 1455 Pennsylvania Avenue, N.W.
; CITY: Washington, D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/258,639A
; FILING DATE: 10 JUNE 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Healey, William J.
; REGISTRATION NUMBER: 36,160
; REFERENCE/DOCKET NUMBER: 104385.116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)942-8400
; TELEFAX: (202)942-8484
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 659 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-258-639A-4

Query Match 81.8%; Score 27; DB 1; Length 659;
Best Local Similarity 83.3%; Pred. No. 6.4e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GEANGN 6
DB 530 GEANGN 535

RESULT 15

US-08-900-951-4
; Sequence 4, Application US/08900951
; Patent No. 5918693
; GENERAL INFORMATION:
; APPLICANT: Su, Hongsheng
; APPLICANT: Blain, Francoise

APPLICANT: Bennett, Clark
APPLICANT: Gu, Kangfu
APPLICANT: Zimmermann, Joseph
APPLICANT: Musil, Roy
TITLE OF INVENTION: Nucleic Acid Sequences And Expression
TITLE OF INVENTION: Systems For Heparinase II And Heparinase III Derived From
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Hale and Dorf
STREET: 1455 Pennsylvania Avenue, N.W.
CITY: Washington, D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC DOS/MS DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/900,951
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/258,639
FILING DATE: 10 JUNE 1994
ATTORNEY/AGENT INFORMATION:
NAME: Healey, William J.
REGISTRATION NUMBER: 36,160
REFERENCE/DOCKET NUMBER: 104385.116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)942 8400
TELEFAX: (202)942 8484
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 659 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-900-951-4

Query Match 81.8%; Score 27; DB 2; Length 659;
Best Local Similarity 83.3%; Pred. No. 6.4e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GEANGN 6
DB 530 GEANGN 535

Search completed: March 6, 2001, 12:53:47
Job time: 347 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:50:45 ; Search time 70.34 Seconds
(without alignments)
5.792 Million cell updates/sec

Title: US-09-196-161D-3

Perfect score: 33

Sequence: 1 GEANGN 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.66:**
2: PIR1:**
3: PIR2:**
4: PIR3:**
5: PIR4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	100.0	395	2 A46031	Immobilization sur
2	33	100.0	642	2 C81345	hypothetical prote
3	30	90.9	292	2 S77139	hypothetical prote
4	30	90.9	468	2 H72230	glutamate synthase
5	30	90.9	588	2 F70971	hypothetical glyci
6	30	90.9	1176	2 A33856	surface-layer 125k
7	29	87.9	217	2 T31830	hypothetical prote
8	29	87.9	344	2 S50474	hypothetical prote
9	29	87.9	473	2 T15986	hypothetical prote
10	29	87.9	476	2 A56510	hypothetical prote
11	29	87.9	477	2 T04916	hypothetical prote
12	29	87.9	583	1 RDZQTB	hypothetical prote
13	29	87.9	608	1 RDZOK1	hypothetical prote
14	29	87.9	622	2 T51223	hypothetical prote
15	29	87.9	623	2 J06568	hypothetical prote
16	29	87.9	791	2 D64629	iron-regulated out
17	29	87.9	791	2 G71885	probable iron-regu
18	29	87.9	974	2 T04910	hypothetical prote
19	29	87.9	1347	2 T02214	ubiquitous TPR mot
20	29	87.9	2145	2 J04747	adenylate cyclase
21	28	84.8	117	1 HVC0C4	ig heavy chain pre
22	28	84.8	205	2 T07959	probable cell wall
23	28	84.8	291	2 G81407	hypothetical prote
24	28	84.8	328	2 G71838	probable nadh oxid
25	28	84.8	332	2 T18923	hypothetical prote
26	28	84.8	335	2 S35542	replication-associ
27	28	84.8	389	2 F83538	chitin-binding pro
28	28	84.8	391	2 T38822	hypothetical prote
29	28	84.8	409	2 S57689	hypothetical prote

30	28	84.8	416	2 S48957	hypothetical prote
31	28	84.8	422	2 S49012	Yolk protein 3 - b
32	28	84.8	443	2 JE0278	GHR receptor homo
33	28	84.8	448	2 I51883	alpha-2B-adrenergic
34	28	84.8	453	2 A35642	alpha-2B-adrenergic
35	28	84.8	455	2 S28221	alpha-2-C2 adrenergic
36	28	84.8	557	2 S62002	hypothetical prote
37	28	84.8	605	2 T05243	oligo-1,6-glucosid
38	28	84.8	718	2 S48726	hypothetical prote
39	28	84.8	806	2 S31479	phenylalanine ammo
40	28	84.8	878	2 A41055	sucrose synthase (
41	28	84.8	972	2 F71608	ecdysone receptor
42	28	84.8	1199	2 S77082	hypothetical prote
43	28	84.8	1271	2 A45555	pyruvate (flavodox
44	28	84.8	1360	2 T18403	glutamate rich pro
45	28	84.8	1360	2 T18403	asparagine/asparta

ALIGNMENTS

RESULT 1

A46031

Immobilization surface I-antigen precursor - Ichthyophthirius multifiliis (fragment)
C:Species: Ichthyophthirius multifiliis

C:Date: 29-Sep-1999 #sequence-revision 29-Sep-1999 #text-change 07-Dec-1999

C:Accession: A46031

R:Clark, T.G.; McGraw, R.A.; Dickerson, H.W.
Proc. Natl. Acad. Sci. U.S.A. 89, 6363-6367, 1992

A:Title: Developmental expression of surface antigen genes in the parasitic ciliate I

A:Reference number: A46031; MUID:92335298

A:Accession: A46031

A:Molecule type: mRNA; protein

A:Residues: 1-395 <CLAS>

A:Cross-references: GB:M92907; NID:93628568; PIDN:AAC36158.1; PID:93628566

F:2-395/Product: Immobilization surface I-antigen #status experimental <MAN>

A:Note: The authors translated the codon UUG for residue 330 as Ile

A:Note: sequence extracted from NCBI backbone (NCBIN:108734, NCBIPI:108735); the sequ

C:Genetics:

A:Genetic code: SGCS

C:Keywords: glycoprotein; surface antigen

F:2-395/Product: Immobilization surface I-antigen #status experimental <MAN>

F:156,191,245,281/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 100.0%; Score 33; DB 2; Length 395;

Best Local Similarity 100.0%; Pred. NO. 11;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEANGN 6

Db 41 GEANGN 46

RESULT 2

C81345

hypothetical protein Cj0736 [Imported] - Campylobacter jejuni (strain NCTC 11168)

C:Species: Campylobacter jejuni

C:Date: 31-Mar-2000 #sequence-revision 31-Mar-2000 #text-change 31-Mar-2000

C:Accession: C81345

R:Parikh, J.; Wren, B.W.; Mungall, K.; Kelley, J.M.; Churcher, C.; Basham, D.; Chai

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals

A:Reference number: A81250; MUID:20150912

A:Accession: C81345

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-642 <PAR>

A:Cross-references: GB:AL139076; GB:AL111168; NID:96968128; PIDN:CAB73010.1; PID:9696

C:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: Cj0736

Query Match 100.0%; Score 33; DB 2; Length 642;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEANGN 6
 | | | | |
 Db 197 GEANGN 202

RESULT 3

S77139
 Hypothetical protein s111757 - *Synechocystis* sp. (strain PCC 6803)

C:Species: *Synechocystis* sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999

C:Accession: S77139
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* S.

A:Reference number: S74322; MUID:97061201

A:Accession: S77139

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-292 <RAN>

A:Cross-references: EMBL:D90908; GB:AB001339; NID:91652725; PIDN:BAAL7697.1; PID:d101843
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 90.9%; Score 30; DB 2; Length 292;
 Best Local Similarity 83.3%; Pred. No. 37;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEANGN 6
 | | | | |
 Db 184 GOANGN 189

RESULT 4

H72230
 glutamate synthase, beta subunit - *Thermotoga maritima* (strain MSB8)

C:Species: *Thermotoga maritima*

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C:Accession: H72230

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Winn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq

A:Reference number: A72200; MUID:99287316

A:Accession: H72230

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-468 <ARN>

A:Cross-references: GB:AE001806; GB:AE000512; NID:94982196; PIDN:AAD36707.1; PID:9498221

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM1640

C:Superfamily: glutamate synthase small chain

Query Match 90.9%; Score 30; DB 2; Length 468;
 Best Local Similarity 83.3%; Pred. No. 61;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEANGN 6
 | | | | |
 Db 345 GDANGN 350

RESULT 5

F70971
 Hypothetical glycine-rich protein RV3367 - *Mycobacterium tuberculosis* (strain H37RV)

C:Species: *Mycobacterium tuberculosis*

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 18-Feb-2000

C:Accession: F70971

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno

A:Reference number: A70500; MUID:98295987

A:Accession: F70971

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA.

A:Residues: 1-588 <COL>

A:Cross-references: GB:AL009198; GB:AL123456; NID:93242262; PIDN:CA15752.1; PID:e120

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV3367

C:Superfamily: unassigned collagens

Query Match 90.9%; Score 30; DB 2; Length 588;
 Best Local Similarity 83.3%; Pred. No. 78;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEANGN 6
 | | | | |
 Db 287 GDANGN 292

RESULT 6

A33856
 Surface-layer 125K protein precursor - *Bacillus sphaericus*

C:Species: *Bacillus sphaericus*

C:Date: 09-Mar-1990 #sequence_revision 09-Mar-1990 #text_change 15-Oct-1999

C:Accession: A33856

R:Bowditch, R.D.; Baumann, P.; Yousten, A.A.

J. Bacteriol. 171, 4178-4188, 1989

A:Title: Cloning and sequencing of the gene encoding a 125-kilodalton surface-layer p

A:Reference number: A33856; MUID:9327128

A:Accession: A33856

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1176 <BOV>

A:Cross-references: GB:M28361; NID:9341911; PIDN:AAA50256.1; PID:9556012

C:Superfamily: S-layer repeat homology

F:93-147/Domain: S-layer repeat homology <SUR2>

Query Match 90.9%; Score 30; DB 2; Length 1176;
 Best Local Similarity 83.3%; Pred. No. 1,6e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEANGN 6
 | | | | |
 Db 58 GDANGN 63

RESULT 7

T31830
 Hypothetical protein C17E7.1 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T31830

R:Bradshaw, R.

submitted to the EMBL Data Library, July 1997

A:Description: The sequence of C. elegans cosmid C17E7.

A:Reference number: Z21091

A:Accession: T31830

A>Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-217 <BRA>
A:Cross-references: EMBL:AF016443; PIDN:AAC24274.1; GSFDB:GN00023; CESP:C17E7.1
A:Experimental source: strain Bristol N2; clone C17E7
C:Genetics:
A:Gene: CESP:C17E7.1
A:Map position: 5
A:Introns: 39/2; 60/2; 80/1

Query Match 87.9%; Score 29; DB 2; Length 217;
Best Local Similarity 83.3%; Pred. No. 45;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEANGN 6
|||
Db 19 GRANGN 24

RESULT 8
S50474
hypothetical protein YER016w - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C>Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 29-Oct-1999
C:Accession: S50474
R:Dieckhoff, F.S.
Submitted to the EMBL Data Library, December 1994
A:Description: The sequence of S. cerevisiae cosmid 9537, 9581, 9495, 9867, and lambda
A:Reference number: S50428
A:Accession: S50474
A:Molecule type: DNA
A:Residues: 1-344 <DLE>
A:Cross-references: EMBL:U018778; NID:9603592; PIDN:AAB64549.1; PID:9603608; MIPS:YER016w
C:Genetics:
A:Gene: SGD:BIM1
A:Cross-references: SGD:S0000818; MIPS:YER016w
A:Map position: 5R

Query Match 87.9%; Score 29; DB 2; Length 344;
Best Local Similarity 83.3%; Pred. No. 74;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GEANGN 6
|||
Db 310 GEANGN 315

RESULT 9
T15986
hypothetical protein F09C12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 24-Nov-1999
C:Accession: T15986
R:Anderson, K.
Submitted to the EMBL Data Library, June 1995
A:Description: The sequence of C. elegans cosmid F09C12.
A:Reference number: Z18442
A:Accession: T15986
A:Status: preliminary; translated from GH/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-473 <AND>
A:Cross-references: EMBL:U28929; NID:9861312; PID:9861317; PIDN:AAA68348.1; CESP:F09C12
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:F09C12.1
A:Introns: 102/3; 132/2; 191/1; 273/2; 344/3; 396/2; 430/2
C:Superfamily: acetylcholine receptor

Query Match 87.9%; Score 29; DB 2; Length 473;
Best Local Similarity 83.3%; Pred. No. 1e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GEANGN 6
|||
Db 431 GETNGN 436

RESULT 10

A56510
dolichyl-diphosphooligosaccharide--protein glycosyltransferase (EC 2.4.1.119) alpha chain
N:Alternate names: N-linked oligosaccharyl transferase, 64k chain; protein J1404; pro
C:Species: Saccharomyces cerevisiae
C>Date: 21-Jul-1995 #sequence_revision 28-Jul-1995 #text_change 29-Oct-1999
C:Accession: A56510; S55186; S56772; S43978; S68778; S68720
R:Silberstein, S.; Collins, P.G.; Kelleher, D.J.; Raplejo, P.J.; Gilmore, R.
J. Cell Biol. 128, 525-536, 1995
A:Title: The alpha subunit of the Saccharomyces cerevisiae oligosaccharyltransferase
A:Reference number: A56510; MUID:95164556
A:Accession: A56510

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-476 <SLD>
A:Cross-references: GR:246719; NID:9683489; PIDN:CAA86674.1; PID:9683490
R:de Haan, M.; Smits, P.H.M.; Grivell, L.A.
Submitted to the EMBL Data Library, May 1995
A:Reference number: S55183
A:Accession: S55186
A:Molecule type: DNA
A:Residues: 1-476 <DEH>

A:Cross-references: EMBL:X87611; NID:9854567; PIDN:CAA60920.1; PID:9854571
R:de Haan, M.; Grivell, L.A.; Smits, P.H.M.
Submitted to the Protein Sequence Database, September 1995
A:Reference number: S56771
A:Accession: S56772
A:Molecule type: DNA

A:Residues: 1-476 <ZAG>
A:Cross-references: EMBL:Z49277; NID:91006707; PIDN:CAA89291.1; PID:91006708; MIPS:YU
R:Knauer, R.; Lehle, L.
FEBS Lett. 344, 83-86, 1994
A:Title: The N-oligosaccharyltransferase complex from yeast.
A:Reference number: S43978; MUID:94237288
A:Accession: S43978

A:Molecule type: protein
A:Residues: 23-29, 'XX', 32-34 <RNA>
R:Patlak, R.; Parker, C.S.; Imperiali, B.
FEBS Lett. 362, 229-234, 1995
A:Title: The essential yeast NLT1 gene encodes the 64 kDa glycoprotein subunit of the
A:Reference number: S68720; MUID:95237386
A:Accession: S68778

A:Molecule type: DNA
A:Residues: 1-476 <PAT>
A:Cross-references: EMBL:U23326; NID:9709966; PIDN:AAA85158.1; PID:9709967
A:Experimental source: strain SEV 6210/6211
A:Accession: S68720

A:Molecule type: protein
A:Residues: 23-42; 283-288, 'N', 290-291, 'T', 293-294; 294-305 <PAW>
A:Experimental source: strain PRY46
C:Genetics:
A:Gene: SGD:OSR1; NLT1
A:Cross-references: SGD:S0003539; MIPS:YUL002C
A:Map position: 10L

C:Function:
A:Description: catalyzes the transfer of core oligosaccharide from dolichol carrier t
A:Pathway: protein glycosylation
C:Keywords: endoplasmic reticulum; glycoprotein; glycosyltransferase; hexosyltransfer
F:1-22/Domin: signal sequence #status predicted <SIG>
F:23-476/Product: oligosaccharyltransferase alpha chain #status experimental <MAT>
F:451-467/Domin: transmembrane #status predicted <TM>
F:18,99,217,336,400,473/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 87.9%; Score 29; DB 2; Length 476;
Best Local Similarity 83.3%; Pred. No. 1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEANGN 6
DB 196 GEGNGN 201

RESULT 11

T04916
hypothetical protein T10114.180 - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 28-Jul-2000
C/Accession: T04916
R/By: M. Murphy, G. Ridley, P. Hudson, S. Bancroft, I. Mewes, H.W. Mayer, K.F.X
submitted to the Protein Sequence Database, April 1998
A/Reference number: Z15389
A/Accession: T04916
A/Molecule type: DNA
A/Residues: 1-477 <BEV>
A/Cross-references: EMBL:AL021712
A/Experimental source: cultivar Columbia; BAC clone T10114
C/Genetics:
A/Map position: 4
A/Intons: 118/3; 200/2; 228/3; 284/3; 330/3; 388/1; 446/2
A/Note: T10114.180
C/Superfamily: Schizosaccharomyces pombe hypothetical protein SPAC577.07

Query Match 87.9%; Score 29; DB 2; Length 477;
Best Local Similarity 83.3%; Pred. No. 1e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GEANGN 6
DB 67 GEGNGN 72

RESULT 12

RD20TB
dihydrofolate reductase (EC 1.5.1.3) / thymidylate synthase (EC 2.1.1.45) - Plasmodium C
C/Species: Plasmodium chabaudi
C/Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 11-Jun-1999
C/Accession: A33484
R/Comman, A.F.; Lew, A.M.
Mol. Cell. Biol. 9, 5182-5188, 1989
A/Title: Antifolate drug selection results in duplication and rearrangement of chromosom
A/Reference number: A33484; MUID:90097935
A/Accession: A33484
A/Molecule type: DNA
A/Residues: 1583 <COM>
A/Cross-references: GB:M30834; NID:g160265; PIDN:AAA29587.1; PID:g160266
C/Superfamily: bifunctional dihydrofolate reductase-thymidylate synthase; thymidylate s
C/Keywords: deoxyribonucleotide biosynthesis; methyltransferase; multifunctional enzym
F:24-164/Domain: type I dihydrofolate reductase homology <DFR>
F:300-583/Domain: thymidylate synthase homology <TDS>
F:465/Active site: Cys #status predicted

Query Match 87.9%; Score 29; DB 1; Length 583;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GEANGN 6
DB 364 GEGNGN 369

RESULT 13

RD2OK1
dihydrofolate reductase (EC 1.5.1.3) / thymidylate synthase (EC 2.1.1.45) - malaria para
C/Species: Plasmodium falciparum
C/Date: 31-Mar-1990 #sequence_revision 31-Mar-1993 #text_change 09-Jun-2000
C/Accession: A39975; A31262; G31262; E31262; F31262; J50208; D31262
R/Beik, D.-J.; Li, W.; Horii, T.; Inselburg, J.
Proc. Natl. Acad. Sci. U.S.A. 84, 8360-8364, 1987

A/Title: Molecular cloning and sequence analysis of the Plasmodium falciparum dihydro
A/Reference number: A39975; MUID:88068594
A/Accession: A39975
A/Molecule type: DNA
A/Residues: 1-608 <BE1>

A/Cross-references: GB:J03028; NID:g160261; PIDN:AAA29585.1; PID:g160262
R/Comman, A.F.; Morry, M.J.; Biggs, B.A.; Cross, G.A.M.; Foote, S.J.
Proc. Natl. Acad. Sci. U.S.A. 85, 9109-9113, 1988

A/Title: Amino acid changes linked to pyrimethamine resistance in the dihydrofolate r
A/Reference number: A94216; MUID:89057885
A/Accession: A31262

A/Molecule type: DNA
A/Residues: 1-608 <COM1>
A/Cross-references: GB:J03772; NID:g340507; PIDN:AAB59212.1; PID:g623444

A/Accession: B31262
A/Molecule type: DNA
A/Residues: 1-50, 'I', 52-608 <COM2>

A/Cross-references: GB:J03772
A/Note: these two sequences are from two different clones, HB3 and 7G8
A/Accession: G31262

A/Molecule type: DNA
A/Residues: 1-15, 'V', 17-107, 'T', 109-608 <COM>
A/Accession: C31262

A/Molecule type: DNA
A/Residues: 1-107, 'S', 109-258 <CO2>
A/Accession: E31262

A/Molecule type: DNA
A/Residues: 1-58, 'R', 60-258 <CO3>
A/Accession: F31262

A/Molecule type: DNA
A/Residues: 1-58, 'R', 60-163, 'L', 165-258 <CO4>
R/Snewin, V.A.; England, S.M.; Sims, P.F.G.; Hyde, J.E.
Gene 76, 41-52, 1989

A/Title: Characterisation of the dihydrofolate reductase-thymidylate synthetase gene
Gene 76, 41-52, 1989
A/Reference number: J50208; MUID:89306658

A/Accession: J50208
A/Molecule type: DNA
A/Residues: 1-58, 'R', 60-608 <SNE>

A/Cross-references: GB:M22159; NID:g160259; PIDN:AAA29580.1; PID:g160260
A/Experimental source: strain K1

A/Note: the authors suggest that translation may begin at Met-2
C/Comment: Dihydrofolate reductase catalyzes the production of methylenetetrahydrofol
C/Genetics:

A/Map position: 4
C/Superfamily: bifunctional dihydrofolate reductase-thymidylate synthase; thymidylat
C/Keywords: deoxyribonucleotide biosynthesis; methyltransferase; multifunctional enzy
F:27-163/Domain: type I dihydrofolate reductase homology <DFR>

F:325-608/Domain: thymidylate synthase homology <TDS>
F:450/Active site: Cys #status predicted

Query Match 87.9%; Score 29; DB 1; Length 608;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GEANGN 6
DB 389 GEGNGN 394

RESULT 14

T51223
hypothetical protein B24M22.190 [imported] - Neurospora crassa
C/Species: Neurospora crassa
C/Date: 28-Jul-2000 #sequence_revision 28-Jul-2000 #text_change 28-Jul-2000
C/Accession: T51223
R/Schulte, U.; Algn, V.; Hohnsels, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu
submitted to the Protein Sequence Database, July 2000
A/Reference number: Z25286
A/Accession: T51223
A/Status: preliminary
A/Molecule type: DNA

A:Residues: 1-622 <SCH>
A:Cross-references: EMBL:AL390354; GSPDB:GN00116; NCSP:B24M22.190
A:Experimental source: strain OR74A
C:Genetics:
A:Gene: NCSP:B24M22.190
A:Map position: 6

Query Match 87.9%; Score 29; DB 2; Length 622;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GEANGN 6
|| |||
Db 607 GEVNGN 612

RESULT 15
JC6568
dihydrofolate reductase (EC 1.5.1.3) / thymidylate synthase (EC 2.1.1.45) - Plasmodium v
C:Species: Plasmodium vivax
C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 04-Jun-1999
C:Accession: JC6568
R:de Becoulas, P.E.; Basco, L.K.; Tahar, R.; Ouatas, T.; Mazabraud, A.
Gene 211, 177-185, 1998
A:Title: Analysis of the Plasmodium vivax dihydrofolate reductase-thymidylate synthase g
A:Reference number: JC6568; MUID:98241515
A:Accession: JC6568
A:Molecule type: DNA
A:Residues: 1-623 <DEP>
A:Cross-references: EMBL:X98123
C:Genetics:
A:Gene: dhfr-ts
C:Superfamily: bifunctional dihydrofolate reductase--thymidylate synthase; thymidylate s
C:Keywords: methyltransferase; NADP; oxidoreductase
F:340-623/Domain: thymidylate synthase homology <TDS>

Query Match 87.9%; Score 29; DB 2; Length 623;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GEANGN 6
|| |||
Db 404 GETNGN 409

Search completed: March 6, 2001, 12:50:48
Job time: 169 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:54:31 ; Search time 38.83 Seconds
(without alignments)
4.990 Million cell updates/sec

Title: US-09-196-161d-3

Perfect score: 33

Sequence: 1 GEANGN 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt.39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DR	ID	Description
1	30	90.9	1176	1	SLAP_BACSH	P38537 bacillus sp
2	29	87.9	344	1	BIM1_YEAST	P40013 saccharomyc
3	29	87.9	476	1	OSTA_YEAST	P41513 saccharomyc
4	29	87.9	583	1	DRTS_PLACH	P20712 plasmodium
5	29	87.9	587	1	DRTS_PLACH	Q27713 plasmodium
6	29	87.9	608	1	DRTS_PLACH	P13922 plasmodium
7	29	87.9	623	1	DRTS_PLACH	Q02604 plasmodium
8	29	87.9	1347	1	UT1_HUMAN	Q14607 homo sapien
9	29	87.9	2145	1	CYAA_PODAN	Q01513 podospira a
10	28	84.8	114	1	RSBV_LISMO	Q05016 listeria mo
11	28	84.8	117	1	HV03_CAICR	Q03982 calman croc
12	28	84.8	382	1	AZAB_DIDMA	Q07715 didelphis m
13	28	84.8	384	1	AZAB_ELEMA	Q19014 elephas max
14	28	84.8	386	1	AZAB_AMBHO	Q18935 amblysomus
15	28	84.8	387	1	AZAB_MACPR	Q19025 macroscelid
16	28	84.8	388	1	AZAB_ORYAF	Q19032 oryctolopus
17	28	84.8	389	1	AZAB_PROHA	Q19034 procavia ca
18	28	84.8	390	1	AZAB_DUGDU	Q07713 dugong dugo
19	28	84.8	409	1	YG4S_YEAST	P50082 saccharomyc
20	28	84.8	416	1	YHRS_YEAST	P38823 saccharomyc
21	28	84.8	448	1	AZAB_CAVPO	Q60475 cavla porce
22	28	84.8	453	1	AZAB_RAT	P19328 rattus norv
23	28	84.8	455	1	AZAB_MOUSE	P30545 mus musculu
24	28	84.8	520	1	CET1_CANAL	Q93803 candida alb
25	28	84.8	557	1	APG6_YEAST	Q02948 saccharomyc
26	28	84.8	718	1	PAL3_PETCR	P45729 petroselinu
27	28	84.8	805	1	SUSY_MEDSA	Q65026 medicago sa
28	28	84.8	806	1	SUSY_VICFA	P31926 vicia faba
29	28	84.8	878	1	ECR_DROME	P34021 drosophilla
30	28	84.8	1199	1	NIFI_SYNY3	P52965 synecocyst
31	28	84.8	1557	1	DVAL_DICVI	Q24702 dictyocaulu
32	28	84.8	1569	1	GLI3_XENLA	Q91660 xenopus lae
33	28	84.8	1638	1	BRM_DROME	P25439 drosophilla

34	27	81.8	152	1	SODC_PRIGL	P11418 prionace gl
35	27	81.8	155	1	YC35_HAEIN	P44131 haemophilus
36	27	81.8	164	1	SIEA_BPP22	Q38673 bacterioph
37	27	81.8	242	1	HAP5_YEAST	Q02516 saccharomyc
38	27	81.8	250	1	MTR3_YEAST	P48240 saccharomyc
39	27	81.8	265	1	Y212_HUMAN	P49751 homo sapien
40	27	81.8	311	1	YG95_MYCLE	Q49897 mycobacteri
41	27	81.8	344	1	HEMA_IHAR	P28730 influenza a
42	27	81.8	344	1	HEMA_IATVA	P12590 influenza a
43	27	81.8	344	1	HEMA_IATVA	P28731 influenza a
44	27	81.8	373	1	YN28_YEAST	P53829 saccharomyc
45	27	81.8	378	1	CSP_PLACH	P08672 plasmodium

ALIGNMENTS

RESULT	ID	SLAP_BACSH	STANDARD	PRT	1176 AA.
AC	P38537	01-OCT-1994 (Rel. 30, Created)			
DT	01-OCT-1994 (Rel. 30, Last sequence update)				
DT	01-FEB-1996 (Rel. 33, Last annotation update)				
DE	SURFACE-LAYER 125 KDA PROTEIN PRECURSOR.				
OS	Bacillus sphaericus.				
OC	Bacteria: Firmicutes; Bacillus/Clostridium group;				
OC	Bacillus/Staphylococcus group; Bacillus.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=2362;				
RX	MEDLINE=89327128; PubMed=266389;				
RA	Bowditch R.D., Baumann P., Yousten A.A.;				
RT	*Cloning and sequencing of the gene encoding a 125-kilodalton				
RT	surface-layer protein from Bacillus sphaericus 2362 and of a related				
RT	cryptic gene.;				
RL	J. Bacteriol. 171:4178-4188(1989).				
CC	-1- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY				
CC	OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA.				
CC	-1- SUBCELLULAR LOCATION: CELL WALL.				
CC	-1- SIMILARITY: CONTAINS 3 S-LAYER HOMOLOGY (SLH) DOMAINS.				
CC	-----				
CC	THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION				
CC	BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -				
CC	THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS				
CC	USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY				
CC	MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL				
CC	ENTITIES REQUIRES A LICENSE AGREEMENT (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL: M28361; AAA50256.1; .				
DR	PIR: A33856; A33856.				
DR	INTERPRO: IPR001119; .				
DR	PFAM: PF00395; SLH. 3.				
DR	PROSITE: PS01072; SLH_DOMAIN. 2.				
FT	SIGNAL: Cell wall; S-layer; Repeat.				
FT	CHAIN 1 30				
FT	POTENTIAL.				
FT	DOMAIN 31 1176				
FT	SLH 1.				
FT	DOMAIN 92 151				
FT	SLH 2.				
FT	DOMAIN 152 210				
FT	SLH 3.				
SQ	SEQUENCE 1176 AA; 125225 MW; IAGVA58EF43788C CRC64;				

Query Match 90.98; Score 30; DB 1; Length 1176;

Best Local Similarity 83.3%; Pred. No. 73;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Query 1 GEANGN 6
Db 58 GDANGN 63

```

RESULT 2
BIM1 YEAST STANDARD; PRT; 344 AA.
AC P40013;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE BIM1 PROTEIN.
GN BIM1 OR YER016W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Bero A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Huntcke-Smith S.,
RA Hyman R., Kayser A., Komp C., Lasker D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Norgren R., Oelner P., Oh C.,
RA Patel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA Taylor P., Wei Y., Yellon M., Botstein D., Davis R.W.;
RA Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RL [2]
RN CHARACTERIZATION.
RP Schwartz K., Richards K., Botstein D.;
RA "BIM1 Encodes a microtubule-binding protein in yeast.";
RL Mol. Biol. Cell 8:2677-2691(1997).
CC -1- FUNCTION: BINDS MICROTUBULES.
CC -1- SIMILARITY: TO S.POMBE MAL3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U18778; AAB64549.1; -
DR SGD; S0000818; BIM1.
DR Microtubules.
KW SEQUENCE 344 AA; 38361 MW; C9087C846A32BD74 CRC64;
SQ

```

```

Query Match 87.9%; Score 29; DB 1; Length 344;
Best Local Similarity 83.3%; Pred. No. 35;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 GEANGN 6
   1 1 1 1 1
Db 310 GEVNGN 315

RESULT 3
OSTA YEAST STANDARD; PRT; 476 AA.
AC P41543;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE DOLICHYL-DIPHOSPHOOLIGOSACCHARIDE--PROTEIN GLYCOSYLTRANSFERASE ALPHA
DE SUBUNIT (OLIGOSACCHARYL TRANSFERASE 64 KDA SUBUNIT).
GN OST1 OR NLT1 OR YJL002C OR J1404.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.; AND SEQUENCE OF 23-38 AND 135-165.
RX MEDLINE=95164556; PubMed=7860628;
RA Silberstein S., Collins P.G., Kelleher D.J., Rapiejko P.J.,
RA Gilmore R.;

```

```

RT "The alpha subunit of the Saccharomyces cerevisiae
RT oligosaccharyltransferase complex is essential for vegetative growth
RT of yeast and is homologous to mammalian ribophorin I.";
RL J. Cell Biol. 128:525-536(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95237386; PubMed=7720878;
RA Pathak R., Parker C.S., Imperial B.;
RT "The essential yeast NLT1 gene encodes the 64 kDa glycoprotein
RT subunit of the oligosaccharyl transferase.";
RL FEBS Lett. 362:229-234(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RA de Haan M., Smits P.H.M., Grievell L.A.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ESSENTIAL SUBUNIT OF N-OLIGOSACCHARYL TRANSFERASE.
CC N-OLIGOSACCHARYL TRANSFERASE CATALYSES THE TRANSFER OF A HIGH
CC MANNOSE OLIGOSACCHARIDE FROM A LIPID-LINKED OLIGOSACCHARIDE DONOR
CC ONTO ASPARAGINE ACCEPTOR SITES WITHIN AN ASN-X-SER/THR CONSENSUS
CC MOTIF IN NEWLY SYNTHESIZED PROTEINS.
CC -1- CATALYTIC ACTIVITY: DOLICHYL DIPHOSPHOOLIGOSACCHARIDE + PROTEIN
CC L-ASPARAGINE -> DOLICHYL DIPHOSPHATE + A GLYCOPROTEIN WITH THE
CC OLIGOSACCHARIDE CHAIN ATTACHED BY GLYCOSYLAMINE LINKAGE TO PROTEIN
CC L-ASPARAGINE.
CC -1- PATHWAY: GLYCOSYLATION.
CC -1- SUBUNIT: YEAST OST SEEMS TO CONSIST OF SIX DIFFERENT SUBUNITS
CC (ALPHA TO ZETA).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ENDOPLASMIC
CC RETICULUM.
CC -----
CC -1- SIMILARITY: TO MAMMALIAN RIBOPHORIN I.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z46719; CAA86674.1; -
DR EMBL; U22326; AAA85158.1; -
DR EMBL; X87611; CAA60920.1; -
DR EMBL; Z49277; CAA89291.1; -
DR SGD; S0003539; OST1.
KW Transferase; Endoplasmic reticulum; Transmembrane; Glycoprotein;
KW Signal.
FT SIGNAL 1 22
FT CHAIN 23 476
FT 1

```

```

DOMAIN 23 449
FT TRANSMEM 450 470
FT DOMAIN 471 476
FT CARBOHYD 18 18
FT CARBOHYD 99 99
FT CARBOHYD 217 217
FT CARBOHYD 336 336
FT CARBOHYD 400 400
SQ SEQUENCE 476 AA; 54072 MW; EFD989B50AAED348 CRC64;

```

```

Query Match 87.9%; Score 29; DB 1; Length 476;
Best Local Similarity 83.3%; Pred. No. 48;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GEANGN 6
   1 1 1 1 1
Db 196 GKANGN 201

RESULT 4
DRTS_PLACH

```

ID DRTS_PLACH STANDARD; PRT: 583 AA.
AC P20712: 027715;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE BIFUNCTIONAL DIHYDROFOLATE REDUCTASE-THYMIDYLATE SYNTHASE (DHER-TS)
DE [INCLUDES: DIHYDROFOLATE REDUCTASE (EC 1.5.1.3); THYMIDYLATE SYNTHASE (EC 2.1.1.45)].
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEGLINE=90097935; PubMed=7601715;
RA Cowman A.F., Lew A.M.;
RT "Antifolate drug selection results in duplication and rearrangement of chromosome 7 in Plasmodium chabaudi.";
RL Mol. Cell. Biol. 9:5182-5186(1989).
[2]
RN SEQUENCE OF 12-193 FROM N.A.
RP MEDLINE=95059225; PubMed=7969277;
RA Cheng O., Saul A.;
RT "The dihydrofolate reductase domain of rodent malarial: point mutations and pyrimethamine resistance.";
RL Mol. Biochem. Parasitol. 65:361-363(1994).
CC -1- CATALYTIC ACTIVITY: 5,6,7,8-TETRAHYDROFOLATE + NADPH =
CC -1- CATALYTIC ACTIVITY: 5,10-METHYLENETHETRAHYDROFOLATE + DUMP =
CC -1- PATHWAY: ESSENTIAL STEP FOR DE NOVO GLYCINE AND PURINE SYNTHESIS,
CC DNA PRECURSOR SYNTHESIS, AND FOR THE CONVERSION OF DUMP TO DTMP.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
CC DIHYDROFOLATE REDUCTASE FAMILY.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE THYMIDYLATE
CC SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M30834; AAA29587.1; -;
DR EMBL: L28120; AAB59201.1; -;
DR PIR: A33484; RDZOTR.
DR HSSP: P00470; 1AN5.
DR INTERPRO: IPR000398; -;
DR INTERPRO: IPR001796; -;
DR PFAM: PF00186; Dihfolate_red; 1.
DR PFAM: PF00303; thymidylat_synt; 1.
DR PRINTS: PR00070; DHER.
DR PRINTS: PR00108; THYMSNTHASE.
DR PROSITE: PS00075; DHER; 1.
DR PROSITE: PS00091; THYMIDYLATE_SYNTHASE; 1.
DR MULTIFUNCTIONAL ENZYME: Oxidoreductase; Transferase; NADP;
DR METHYLTRANSFERASE; Nucleotide biosynthesis; One-carbon metabolism.
KW Methylintransferase; Nucleotide biosynthesis; One-carbon metabolism.
FT DOMAIN 1 232 DIHYDROFOLATE REDUCTASE.
FT ACT_SITE 298 583 THYMIDYLATE SYNTHASE.
FT VARIANT 465 465 BY SIMILARITY.
FT VARIANT 106 106 S -> I (IN PYRIMETHAMINE RESISTANCE).
FT CONFLICT 27 27 G -> S (IN REF. 2).
FT CONFLICT 156 156 C -> S (IN REF. 2).
FT CONFLICT 164 164 I -> V (IN REF. 2).
SQ SEQUENCE 583 AA; 68051 MW; 4AA55E1C987E6ED7 CRC64;

Query Match 87.9%; Score 29; DB 1; Length 583;
Best Local Similarity 83.3%; Pred. No. 59;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
1 GEANGN 6
11111

Db 364 GETNGN 369
RESULT 5
DRTS_PLABA STANDARD; PRT: 587 AA.
AC 027713: 027714;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE BIFUNCTIONAL DIHYDROFOLATE REDUCTASE-THYMIDYLATE SYNTHASE (DHER-TS)
DE [INCLUDES: DIHYDROFOLATE REDUCTASE (EC 1.5.1.3); THYMIDYLATE SYNTHASE (EC 2.1.1.45)].
OS Plasmodium berghei (strain Anka).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95198769; PubMed=7891743;
RA van Dijk M.R., McConkey G.A., Vinkenoog R., Waters A.P., Janse C.J.;
RT "Mechanisms of pyrimethamine resistance in two different strains of Plasmodium berghei.";
RL Mol. Biochem. Parasitol. 68:167-171(1994).
[2]
RN SEQUENCE OF 12-201 FROM N.A.
RP MEDLINE=95059225; PubMed=7969277;
RA Cheng O., Saul A.;
RT "The dihydrofolate reductase domain of rodent malarial: point mutations and pyrimethamine resistance.";
RL Mol. Biochem. Parasitol. 65:361-363(1994).
CC -1- CATALYTIC ACTIVITY: 5,6,7,8-TETRAHYDROFOLATE + NADPH =
CC -1- CATALYTIC ACTIVITY: 5,10-METHYLENETHETRAHYDROFOLATE + DUMP =
CC -1- PATHWAY: ESSENTIAL STEP FOR DE NOVO GLYCINE AND PURINE SYNTHESIS,
CC DNA PRECURSOR SYNTHESIS, AND FOR THE CONVERSION OF DUMP TO DTMP.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
CC DIHYDROFOLATE REDUCTASE FAMILY.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE THYMIDYLATE
CC SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U12275; AAB60237.1; -;
DR EMBL: L28119; AAA29581.1; -;
DR HSSP: P00470; 1AN5.
DR INTERPRO: IPR000398; -;
DR INTERPRO: IPR001796; -;
DR PFAM: PF00186; Dihfolate_red; 1.
DR PFAM: PF00303; thymidylat_synt; 1.
DR PRINTS: PR00070; DHER.
DR PRINTS: PR00108; THYMSNTHASE.
DR PROSITE: PS00075; DHER; 1.
DR PROSITE: PS00091; THYMIDYLATE_SYNTHASE; 1.
DR MULTIFUNCTIONAL ENZYME: Oxidoreductase; Transferase; NADP;
DR METHYLTRANSFERASE; Nucleotide biosynthesis; One-carbon metabolism.
KW Methylintransferase; Nucleotide biosynthesis; One-carbon metabolism.
FT DOMAIN 1 240 DIHYDROFOLATE REDUCTASE.
FT ACT_SITE 301 587 THYMIDYLATE SYNTHASE.
FT VARIANT 469 469 BY SIMILARITY.
FT VARIANT 110 110 S -> N (IN PYRIMETHAMINE RESISTANCE).
FT VARIANT 177 177 S -> F (IN PYRIMETHAMINE RESISTANCE).
SQ SEQUENCE 587 AA; 68932 MW; 6E36C2B02FEC13A CRC64;

Query Match 87.9%; Score 29; DB 1; Length 587;
Best Local Similarity 83.3%; Pred. No. 60;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GEANGN 6 .
 DB 368 GETNGN 373

RESULT 6

DRTS_PLAFK STANDARD; PRT; 608 AA.
 ID DRTS_PLAFK
 AC P13922; 027734;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE BIFUNCTIONAL DIHYDROFOLATE REDUCTASE-THYMIDYLATE SYNTHASE (DHFR-TS)
 DE [INCLUDES: DIHYDROFOLATE REDUCTASE (EC 1.5.1.3); THYMIDYLATE SYNTHASE (EC 2.1.1.45)]
 OS Plasmodium falciparum (isolate K1 / Thailand).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ISOLATE K1, AND ISOLATE FCR3;
 RX MEDLINE-8930658; PubMed-2663650;
 RA Snewin V.A., England S.M., Sims P.F.G., Hyde J.E.;
 RT "Characterisation of the dihydrofolate reductase-thymidylate synthetase gene from human malaria parasites highly resistant to pyrimethamine";
 RL Gene 76:41-52(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-90014979; PubMed-2677719;
 RA Zolg J.W., Plitt J.R., Chen G.-X., Palmer S.;
 RT "Point mutations in the dihydrofolate reductase-thymidylate synthase gene as the molecular basis for pyrimethamine resistance in Plasmodium falciparum";
 RL Mol. Biochem. Parasitol. 36:253-262(1989).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ISOLATE PALO-ALTO;
 RX MEDLINE-89057885; PubMed-3057499;
 RA Corman A.F., Morry M.J., Biggs B.A., Cross G.A.M., Foote S.J.;
 RT "Amino acid changes linked to pyrimethamine resistance in the dihydrofolate reductase-thymidylate synthase gene of Plasmodium falciparum";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:9109-9113(1988).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ISOLATE FCR3;
 RX MEDLINE-88068594; PubMed-2825189;
 RA Baik D.J., Li W.B., Horii T., Inselburg J.;
 RT "Molecular cloning and sequence analysis of the Plasmodium falciparum dihydrofolate reductase-thymidylate synthase gene";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:8360-8364(1987).
 CC -1- CATALYTIC ACTIVITY: 5,6,7,8-TETRAHYDROFOLATE + NADPH.
 CC -1- CATALYTIC ACTIVITY: 5,10-METHYLENETHETRAHYDROFOLATE + DUMP = DIHYDROFOLATE + DUMP.
 CC -1- PATHWAY: ESSENTIAL STEP FOR DE NOVO GLYCINE AND PURINE SYNTHESIS, DNA PRECURSOR SYNTHESIS, AND FOR THE CONVERSION OF DUMP TO DUMP.
 CC -1- MISCELLANEOUS: K1 IS FROM A PYRIMETHAMINE-RESISTANT STRAIN, FCR3 IS A PYRIMETHAMINE-SENSITIVE STRAIN.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE DIHYDROFOLATE REDUCTASE FAMILY.
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE THYMIDYLATE SYNTHASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M22159; AAA29580.1; .

DR EMBL: J04643; AAA29586.1; .
 DR EMBL: J03772; AAB59212.1; .
 DR EMBL: J03028; AAA29585.1; .
 DR PIR: A39975; RDZOK1.
 DR PIR: G31262; G31262.
 DR HSSP: P00470; .IANS.
 DR INTERPRO: IPR000398; .
 DR INTERPRO: IPR001796; .
 DR PFAM: PF00186; Dihfolate_red; 1.
 DR PFAM: PF00303; Thymidylat_synth; 1.
 DR PRINTS: PR00070; DHFR.
 DR PRINTS: PR00108; THYMSNTASE.
 DR PROSITE: PS00075; DHFR; 1.
 DR PROSITE: PS00091; THYMIDYLATE_SYNTHASE; 1.
 DR MULTIFUNCTIONAL ENZYME: Oxidoreductase; Transferase; NADP;
 KW Methylenetetrahydrofolate reductase; Dihydrofolate reductase; Nucleotide biosynthesis; One-carbon metabolism.
 FT DOMAIN 1 231
 FT ACT_SITE 322 608
 FT ACT_SITE 490 490
 FT VARIANT 16 16
 FT VARIANT 51 51
 FT VARIANT 59 59
 FT VARIANT 59 59
 FT VARIANT 108 108
 FT VARIANT 108 108
 FT SEQUENCE 608 AA; 71817 MW; 7727EBBA3946996 CRC64;

Query Match 87.9%; Score 29; DB 1; Length 608;
 Best Local Similarity 83.3%; Pred. No. 62;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GEANGN 6
 DB 389 GETNGN 394

RESULT 7
 DRTS_PLAVI STANDARD; PRT; 623 AA.
 ID DRTS_PLAVI
 AC 002604; 015873;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE BIFUNCTIONAL DIHYDROFOLATE REDUCTASE-THYMIDYLATE SYNTHASE (DHFR-TS)
 DE [INCLUDES: DIHYDROFOLATE REDUCTASE (EC 1.5.1.3); THYMIDYLATE SYNTHASE (EC 2.1.1.45)]
 OS Plasmodium vivax.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ISOLATE ARI/PAKISTAN, ISOLATE BUR-98, AND ISOLATE BUR-151;
 RX MEDLINE-98241515; PubMed-9573357;
 RA Eldin de Pecoulas P., Basco L.K., Tahar R., Ouates T., Mazabraud A.;
 RT "Analysis of the Plasmodium vivax dihydrofolate reductase-thymidylate synthase gene sequence";
 RL Gene 211:177-185(1998).
 CC -1- CATALYTIC ACTIVITY: 5,6,7,8-TETRAHYDROFOLATE + NADPH.
 CC -1- CATALYTIC ACTIVITY: 5,10-METHYLENETHETRAHYDROFOLATE + DUMP = DIHYDROFOLATE + DUMP.
 CC -1- PATHWAY: ESSENTIAL STEP FOR DE NOVO GLYCINE AND PURINE SYNTHESIS, DNA PRECURSOR SYNTHESIS, AND FOR THE CONVERSION OF DUMP TO DUMP.
 CC -1- DOMAIN: THE REPEAT REGION IS MISSING IN THE PYRIMETHAMINE-RESISTANT ISOLATES BUR-98 AND BUR-151.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE DIHYDROFOLATE REDUCTASE FAMILY.
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE THYMIDYLATE SYNTHASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its

CC use, by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X98123; CA66805.1; -
 DR HSSP; P00470; IAN5.
 DR INTERPRO; IPR000398; -
 DR INTERPRO; IPR001796; -
 DR PFAM; PF00186; Dihydrate_red; 1.
 DR PFAM; PF00303; thymidylat_synt; 1.
 DR PRINTS; PR00070; DHFR.
 DR PRINTS; PR00108; THYMDSNTASE.
 DR PROSITE; PS00075; DHFR; 1.
 DR PROSITE; PS00091; THYMIDYLATE SYNTHASE; 1.
 KM Multifunctional enzyme; Oxidoreductase; Transferrase; NADP; Repeat;
 FT Methylotransferrase; Nucleotide biosynthesis; One-carbon metabolism.
 FT DOMAIN 1 240 DIHYDROFOLATE REDUCTASE.
 FT DOMAIN 337 623 THYMIDYLATE SYNTHASE.
 FT DOMAIN 88 103 3 X 4 AA REPEATS OF G-G-D-N.
 FT REPEAT 88 91 1.
 FT REPEAT 94 97 2.
 FT REPEAT 100 103 3.
 FT ACT_SITE 505 505
 FT VARIANT 58 58
 FT VARIANT 117 117
 FT VARIANT S -> R (IN THE PYRIMETHAMINE-RESISTANT
 FT ISOLATES BUR-98 AND BUR-151).
 FT S -> N (IN THE PYRIMETHAMINE-RESISTANT
 FT ISOLATES BUR-98 AND BUR-151).
 SQ SEQUENCE 623 AA; 71056 MW; 3E6E958F04FB5828 CRC64;

Query Match 87.9%; Score 29; DB 1; Length 623;
 Best Local Similarity 83.3%; Pred. No. 63;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GEANGN 6
 111111

DB 404 GETNGN 409

RESULT 8
 UTY_HUMAN STANDARD; PRT; 1347 AA.
 AC 014607; 014608;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE UBIQUITOUSLY TRANSCRIBED Y CHROMOSOME TETRAPEPTIDE REPEAT PROTEIN
 DE (UBIQUITOUSLY TRANSCRIBED TPR PROTEIN ON THE Y CHROMOSOME).
 GN UTY.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE;98022381; PubMed-9381176;
 RA Lahn B.T.; Page D.C.;
 RT "Functional coherence of the human Y chromosome.";
 RL Science 278:675-680(1997).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST TWO ISOFORMS ARE PRODUCED BY
 CC ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: CONTAINS 5 TPR DOMAINS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AF000994; AAC51841.1; -

DR EMBL; AF000995; AAC51842.1; -
 DR EMBL; AF000996; AAC51843.1; -
 DR MIM; 400009; -
 DR INTERPRO; IPR001440; -
 DR PFAM; PF00515; TPR; 5.
 KW Repeat; TPR domain; Nuclear protein; Alternative splicing.
 FT DOMAIN 93 121
 FT DOMAIN 130 158
 FT DOMAIN 167 196
 FT DOMAIN 318 346
 FT DOMAIN 352 380
 FT VARSPLIC 996 1079
 FT EENKPTQHKDHSNENSTSSNGRRRKPRTKFGTINID
 FT LSDNKKKKLQJLHETLKLPAFAVAVSAGMLTHVCHTIIIGAN
 FT TV -> AGMWCDSLSLOPPGPKRRESHLSPNSMNYRHL
 FT PSCPTNCFIEVETGFHHYGACLETITSGGLASASQSAGI
 FT TGVSHHAR (TN SHORT ISOFORM).
 FT TGVSHHAR (TN SHORT ISOFORM).
 SQ SEQUENCE 1347 AA; 149577 MW; C26B870127107E71 CRC64;

Query Match 87.9%; Score 29; DB 1; Length 1347;
 Best Local Similarity 83.3%; Pred. No. 1.4e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEANGN 6
 111111

DB 733 GKANGN 738

RESULT 9
 CYAA_PODAN STANDARD; PRT; 2145 AA.
 ID CYAA_PODAN
 AC 001513;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE ADENYLATE CYCLASE (EC 4.6.1.1) (ATP PYROPHOSPHATE-LYASE) (ADENYLYL
 DE CYCLASE).
 OS Podospora anserina.
 CC Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae; Podospora.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE;96200867; PubMed-8621071;
 RA Loubraudon G.; Begueret J.; Turcy B.;
 RT "An additional copy of the adenylate cyclase-encoding gene relieves
 RT developmental defects produced by a mutation in a vegetative
 RT incompatibility-controlling gene in Podospora anserina.";
 RL Gene 170:119-123(1996).
 CC -1- FUNCTION: PLAYS ESSENTIAL ROLES IN REGULATION OF CELLULAR
 CC METABOLISM BY CATALYSING THE SYNTHESIS OF A SECOND MESSENGER,
 CC CAMP.
 CC -1- CATALYTIC ACTIVITY: ATP = 3',5'-CYCLIC AMP + PYROPHOSPHATE.
 CC -1- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
 CC MANY PROTEINS.
 CC -1- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-3 FAMILY.
 CC -1- SIMILARITY: CONTAINS A PP2C-LIKE DOMAIN.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; LA3413; AB05642.1; -
 DR INTERPRO; IPR001611; -
 DR INTERPRO; IPR001932; -
 DR PFAM; PF00560; LRR; 13.
 DR PFAM; PF00481; PP2C; 1.
 DR PRINTS; PR00019; LEURICHPRT.
 DR PROSITE; PS50125; GUANTALATE_CYCLASES_2; 1.
 KW Lyase; Repeat; Leucine-repeat; CAMP synthesis; Magnesium.
 FT DOMAIN 36 41
 FT DOMAIN POLY-SER.

FT DOMAIN 1445 1710 PP2C-LIKE.
 DT DOMAIN 1711 2145 CATALYTIC.
 SQ SEQUENCE 2145 AA; 237515 MW; 88E7EF6E4AC0687D CRC64;

Query Match 87.9%; Score 29; DB 1; Length 2145;
 Best Local Similarity 83.3%; Pred. No. 2, 2e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GEANGN 6
 11111
 Db 1125 GEINGN 1130

RESULT 10
 RSBV_LISMO
 ID RSBV_LISMO STANDARD; PRT; 114 AA.
 AC 085016;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE 30-MAY-2000 (Rel. 39, Last annotation update)
 DE ANTI-SIGMA B FACTOR ANTAGONIST.
 GN RSBV
 OS Listeria monocytogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Listeria.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-689426;
 RX MEDLINE-98324975; PubMed-9658010;
 RA Wiedmann M., Arvik T.J., Hurley R.J., Boor K.J.;
 RT "General stress transcription factor sigmaB and its role in acid
 tolerance and virulence of Listeria monocytogenes."
 RL J. Bacteriol. 180:3650-3656(1998).

CC -1- FUNCTION: COUNTERACTS THE RSBV-MEDIATED INHIBITION OF SIGMA B BY
 BINDING DIRECTLY TO RSBV AND BLOCKING ITS ABILITY TO FORM THE
 RSBV-SIGMA B COMPLEX. THIS ACTIVITY OF RSBV APPEARS TO BE
 INHIBITED BY RSBV-DEPENDENT PHOSPHORYLATION (BY SIMILARITY).
 CC -1- PPM: PHOSPHORYLATED BY RSBV ON A SERINE RESIDUE (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE RSBV / SPOIIA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL: AF032444; AAC38787.1; -;
 DR HSSP: P10727; 1B0Z.
 DR INTERPRO: IPR002645; -;
 DR PRAM: PF01740; SpoIIA; 1.
 KW Phosphorylation.
 FT MOD_RES 58
 SQ SEQUENCE 114 AA; 12798 MW; AFCEEE64C146C023 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 114;
 Best Local Similarity 83.3%; Pred. No. 18;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GEANGN 6
 11111
 Db 107 GEINGN 112

RESULT 11
 HV03_CAICR STANDARD; PRT; 117 AA.
 AC P03982;
 DT 23-OCT-1986 (Rel. 02, Created)
 DT 23-OCT-1986 (Rel. 02, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE IG HEAVY CHAIN V REGION G4 PRECURSOR.
 GN G4.

OS Calman crocodilus (Spectacled caiman) (Caiman sclerops).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Crocodylidae; Alligatorinae; Caiman.
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE-85140192; PubMed-2983316;
 RA Litman G.W., Murphy K., Berger L., Litman R., Hinds K.,
 RA Erickson B.W.;
 RT "Complete nucleotide sequences of three VH genes in Caiman, a
 phylogenetically ancient reptile: evolutionary diversification in
 coding segments and variation in the structure and organization of
 recombination elements."
 RT Proc. Natl. Acad. Sci. U.S.A. 82:844-848(1985).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL: M12770; AAA49194.1; -;
 DR PIR: A02085; HVC0G4.
 DR INTERPRO: IPR003006; -;
 DR PRAM: PF00047; 1g; 1.
 KW Immunoglobulin V region; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 117 IG HEAVY CHAIN V REGION G4.
 FT DOMAIN 20 49 FRAMEWORK 1.
 FT DOMAIN 50 54 COMPLEMENTARITY-DETERMINING 1.
 FT DOMAIN 55 68 FRAMEWORK 2.
 FT DOMAIN 69 85 COMPLEMENTARITY-DETERMINING 2.
 FT DOMAIN 86 117 FRAMEWORK 3.
 FT DISULFID 41 115 BY SIMILARITY.
 FT NON_TER 117 117
 SQ SEQUENCE 117 AA; 12994 MW; 6330D7469AAE55FC CRC64;

Query Match 84.8%; Score 28; DB 1; Length 117;
 Best Local Similarity 83.3%; Pred. No. 19;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GEANGN 6
 11111
 Db 68 GEINGN 73

RESULT 12
 A2AB_DIDMA STANDARD; PRT; 382 AA.
 ID A2AB_DIDMA
 AC 077715;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ALPHA-2B ADRENERGIC RECEPTOR (ALPHA-2B ADRENERGIC RECEPTOR) (FRAGMENT).
 GN ADR2B.
 OS Didelphis marsupialis virginiana (North American opossum).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Metatheria; Didelphimorphia; Didelphidae; Didelphis.
 RN [1]
 RP SEQUENCE FROM N.A.

RA Stanhope M.T., Madsen O., Madsen V.G., Cleven G.C., de Jong W.W.,
 RA Springer M.S.;
 RT "Highly congruent molecular support for a diverse superordinal clade
 of edemic African mammals."
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: ALPHA-2 ADRENERGIC RECEPTORS MEDIATE THE CATECHOLAMINE-
 INDUCED INHIBITION OF ADENYLYLATE CYCLASE THROUGH THE ACTION OF G
 PROTEINS.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y15943; CAA75896.1; -
DR INTERPRO: IPR000207; -
DR INTERPRO: IPR000276; -
DR INTERPRO: IPR002233; -
DR PFAM: PF00001; 7tm_1; 1.
DR PRINTS: PR00237; GPCRHOOPSPN.
DR PRINTS: PR00559; ADRENRCGA2BR.
DR PRINTS: PR01103; ADRENRCGICR.
DR PROSITE: PS00237; G-PROTEIN_RECEP_FL_1; 1.
DR PROSITE: PS50262; G-PROTEIN_RECEP_FL_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT NON_TER 1
FT TRANSMEM <1 25 1 (POTENTIAL).
FT DOMAIN 26 36 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 37 62 2 (POTENTIAL).
FT DOMAIN 63 72 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 73 95 3 (POTENTIAL).
FT DOMAIN 96 117 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 118 140 4 (POTENTIAL).
FT DOMAIN 141 156 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 157 180 5 (POTENTIAL).
FT DOMAIN 181 346 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 347 370 6 (POTENTIAL).
FT DOMAIN 371 379 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 380 >382 7 (POTENTIAL).
FT DISULFID 72 151 BY SIMILARITY.
FT DOMAIN 271 283 ASP/GLU-RICH (ACIDIC).
FT NON_TER 382 382
SQ SEQUENCE 382 AA; 41870 MW; EAF12DD44B2AA19 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 382;
Best Local Similarity 83.3%; Pred. No. 64;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEANGN 6
DB 228 GEANGH 233

RESULT 13
A2AB_ELEMA STANDARD; PRT; 384 AA.
ID A2AB_ELEMA
AC 019014;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ALPHA-2B ADRENERGIC RECEPTOR (ALPHA-2B ADRENOCEPTOR) (FRAGMENT).
GN ADRA2B.
OS Elephas maximus (Indian elephant).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Proboscidea; Elephantiidae; Elephas.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97357151; Pubmed=9214502;
RA Springer M.S., Cleven G.C., Madsen O., de Jong W.W., Waddell V.G.,
RT Amrine H.M., Stanhope M.J.;
RL "Endemic African mammals shake the phylogenetic tree."
RT Nature 388:61-64(1997).
CC -1- FUNCTION: ALPHA-2 ADRENERGIC RECEPTORS MEDIANE THE CATECHOLAMINE-
CC INDUCED INHIBITION OF ADENYLYATE CYCLASE THROUGH THE ACTION OF G
CC PROTEINS.
CC -----

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y12525; CAA73125.1; -
DR GCRDB: GCR_2403; -
DR INTERPRO: IPR000276; -
DR PFAM: PF00001; 7tm_1; 1.
DR PROSITE: PS00237; G-PROTEIN_RECEP_FL_1; 1.
DR PROSITE: PS50262; G-PROTEIN_RECEP_FL_2; 1.
KW G-protein coupled receptor; Transmembrane; Multigene family;
KW Phosphorylation; Lipoprotein; Palmitate.
FT NON_TER 1
FT TRANSMEM <1 25 1 (POTENTIAL).
FT DOMAIN 26 36 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 37 62 2 (POTENTIAL).
FT DOMAIN 63 72 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 73 95 3 (POTENTIAL).
FT DOMAIN 96 117 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 118 140 4 (POTENTIAL).
FT DOMAIN 141 156 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 157 180 5 (POTENTIAL).
FT DOMAIN 181 348 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 349 372 6 (POTENTIAL).
FT DOMAIN 373 381 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 382 >384 7 (POTENTIAL).
FT DISULFID 72 151 BY SIMILARITY.
FT DOMAIN 281 285 ASP/GLU-RICH (ACIDIC).
FT NON_TER 384 384
SQ SEQUENCE 384 AA; 41911 MW; CP41B56CC35B94F CRC64;

Query Match 84.8%; Score 28; DB 1; Length 384;
Best Local Similarity 83.3%; Pred. No. 64;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEANGN 6
DB 232 GEANGH 237

RESULT 14
A2AB_AMBHO STANDARD; PRT; 386 AA.
ID A2AB_AMBHO
AC 018935;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE ALPHA-2B ADRENERGIC RECEPTOR (ALPHA-2B ADRENOCEPTOR) (FRAGMENT).
GN ADRA2B.
OS Amblysomus hottentotus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Insectivora; Chrysochloridae; Amblysomus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97357151; Pubmed=9214502;
RA Springer M.S., Cleven G.C., Madsen O., de Jong W.W., Waddell V.G.,
RT Amrine H.M., Stanhope M.J.;
RL "Endemic African mammals shake the phylogenetic tree."
RT Nature 388:61-64(1997).
CC -1- FUNCTION: ALPHA-2 ADRENERGIC RECEPTORS MEDIANE THE CATECHOLAMINE-
CC INDUCED INHIBITION OF ADENYLYATE CYCLASE THROUGH THE ACTION OF G
CC PROTEINS.
CC -----
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: Y12526; CAA73126.1; -
 DR GCRDB: GCR_2400; -
 DR INTERPRO: IPR000276; -
 DR PFAM: PF00001; 7tm_1; 1.
 DR PROSITE: PS00237; G-PROTEIN_RECEP_F1_1; 1.
 DR PROSITE: PS50262; G-PROTEIN_RECEP_F1_2; 1.
 DR G-protein coupled receptor; Transmembrane; Multigene family;
 KM Phosphorylation; Lipoprotein; Palmitate.
 FT NON_TER 1
 FT TRANSMEM <1 25 1 (POTENTIAL).
 FT DOMAIN 26 36 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 37 62 2 (POTENTIAL).
 FT DOMAIN 63 72 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 73 95 3 (POTENTIAL).
 FT DOMAIN 96 117 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 118 140 4 (POTENTIAL).
 FT DOMAIN 141 156 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 157 180 5 (POTENTIAL).
 FT DOMAIN 181 350 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 351 374 6 (POTENTIAL).
 FT DOMAIN 375 383 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 384 >386 7 (POTENTIAL).
 FT DOMAIN 280 289 ASP/Glu-RICH (ACIDIC).
 FT DISULFID 72 151 BY SIMILARITY.
 FT NON_TER 386 386
 SQ SEQUENCE 386 AA; 42173 MW; F2F0E14215A85F62 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 386;
 Best Local Similarity 83.3%; Pred. No. 64;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEANGN 6
 |||||:
 Db 232 GEANGH 237

RESULT 15
 AZAB_MACPR STANDARD; PRT; 387 AA.
 ID A2AB_MACPR
 AC 019025;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE ALPHA-2B ADRENERGIC RECEPTOR (ALPHA-2B ADRENOCEPTOR) (FRAGMENT).
 GN ADRA2B.
 OS Macroscelidés proboscideus (Short-eared elephant shrew).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Eulestomii;
 OC Mammalia; Eutheria; Macroscelidea; Macroscelididae; Macroscelides.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-97357151; PubMed-9214502;
 RA Springer M.S., Clevén G.C., Madsen O., de Jong W.W., Waddell V.G.,
 RA Amrine H.M., Stanhope M.J.;
 RT "Endemic African mammals shake the phylogenetic tree."
 RL Nature 388:61-64(1997).
 CC -!- FUNCTION: ALPHA-2 ADRENERGIC RECEPTORS MEDIATE THE CATECHOLAMINE-
 CC INDUCED INHIBITION OF ADENYLATE CYCLASE THROUGH THE ACTION OF G
 CC PROTEINS.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: Y12524; CAA73124.1; -
 DR GCRDB: GCR_2404; -
 DR INTERPRO: IPR000276; -
 DR PFAM: PF00001; 7tm_1; 1.
 DR PROSITE: PS00237; G-PROTEIN_RECEP_F1_1; 1.
 DR PROSITE: PS50262; G-PROTEIN_RECEP_F1_2; 1.
 DR G-protein coupled receptor; Transmembrane; Multigene family;
 KM Phosphorylation; Lipoprotein; Palmitate.
 FT NON_TER 1
 FT TRANSMEM <1 25 1 (POTENTIAL).
 FT DOMAIN 26 36 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 37 62 2 (POTENTIAL).
 FT DOMAIN 63 72 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 73 95 3 (POTENTIAL).
 FT DOMAIN 96 117 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 118 140 4 (POTENTIAL).
 FT DOMAIN 141 156 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 157 180 5 (POTENTIAL).
 FT DOMAIN 181 351 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 352 375 6 (POTENTIAL).
 FT DOMAIN 376 384 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 385 >387 7 (POTENTIAL).
 FT DOMAIN 280 288 ASP/Glu-RICH (ACIDIC).
 FT DISULFID 72 151 BY SIMILARITY.
 FT NON_TER 387 387
 SQ SEQUENCE 387 AA; 42488 MW; 7FA95C5B37700E24 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 387;
 Best Local Similarity 83.3%; Pred. No. 65;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEANGN 6
 |||||:
 Db 233 GEANGH 238

Search completed: March 6, 2001, 12:54:33
 Job time: 390 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:52:46 ; Search time 116.78 Seconds
(without alignments)
6.022 Million cell updates/sec

Title: US-09-196-161d-3

Perfect score: 33

Sequence: 1 GEANGN 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_15:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	100.0	298	5 Q9VA15	Q9VA15 dirosophila
2	33	100.0	394	5 Q27208	Q27208 ichtyophth
3	33	100.0	419	5 Q9NHC8	Q9NHC8 branchiosto
4	33	100.0	442	5 Q9XG2	Q9XG2 ichtyophth
5	33	100.0	642	2 Q9PH1	Q9PH1 campylobact
6	30	90.9	292	2 P73652	P73652 synchocyst
7	30	90.9	401	4 Q92554	Q92554 homo sapien
8	30	90.9	468	2 Q9X1X5	Q9X1X5 thermotoga
9	30	90.9	588	2 Q50396	Q50396 mycobacteri
10	30	90.9	720	5 Q9VHC3	Q9VHC3 dirosophila
11	30	90.9	931	5 Q9N9R8	Q9N9R8 leishmania
12	29	87.9	184	10 Q9SUX7	Q9SUX7 arabiidopsis
13	29	87.9	217	5 Q16398	Q16398 caenorhabdi
14	29	87.9	391	5 Q96716	Q96716 branchiosto
15	29	87.9	473	5 Q10052	Q10052 caenorhabdi
16	29	87.9	477	10 Q49640	Q49640 arabiidopsis
17	29	87.9	539	5 Q9N8G3	Q9N8G3 trypanosoma
18	29	87.9	622	3 Q9P389	Q9P389 neurospora
19	29	87.9	707	2 Q50198	Q50198 helicobacte

20	29	87.9	791	2 Q25543	Q25543 helicobacte
21	29	87.9	791	2 Q92KX4	Q92KX4 helicobacte
22	29	87.9	974	10 Q49634	Q49634 arabiidopsis
23	29	87.9	1047	2 Q9RB35	Q9RB35 cytophaga s
24	29	87.9	1399	5 Q9W427	Q9W427 dirosophila
25	28	84.8	165	5 Q9V8N4	Q9V8N4 dirosophila
26	28	84.8	177	2 Q52768	Q52768 rickettsia
27	28	84.8	197	2 Q52765	Q52765 rickettsia
28	28	84.8	205	10 Q22432	Q22432 pinus taeda
29	28	84.8	236	5 Q9V7N1	Q9V7N1 dirosophila
30	28	84.8	265	5 Q17911	Q17911 caenorhabdi
31	28	84.8	291	2 Q9PHR6	Q9PHR6 campylobact
32	28	84.8	307	5 Q9NFE0	Q9NFE0 dirosophila
33	28	84.8	328	2 Q9ZJW2	Q9ZJW2 helicobacte
34	28	84.8	331	2 Q9ZEJ5	Q9ZEJ5 anabena sp
35	28	84.8	332	5 Q17630	Q17630 caenorhabdi
36	28	84.8	333	2 Q9ZEK1	Q9ZEK1 anabena va
37	28	84.8	335	2 Q51587	Q51587 plectonema
38	28	84.8	348	2 Q9K9H4	Q9K9H4 bacillus ha
39	28	84.8	374	5 Q9VDR6	Q9VDR6 dirosophila
40	28	84.8	376	5 Q76880	Q76880 dirosophila
41	28	84.8	376	5 Q9W4X2	Q9W4X2 dirosophila
42	28	84.8	389	2 Q9RM14	Q9RM14 pseudomonas
43	28	84.8	391	3 Q36030	Q36030 schizosacch
44	28	84.8	393	11 Q9JW2	Q9JW2 cavia porce
45	28	84.8	394	2 Q9KY56	Q9KY56 streptomyce

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	298 AA.
Q9VA15	ID Q9VA15			
AC Q9VA15	01-MAY-2000 (TREMBL)	13, Created)		
DT	01-MAY-2000 (TREMBL)	13, Last sequence update)		
DT	01-JUN-2000 (TREMBL)	14, Last annotation update)		
DE	CG11317 PROTEIN.			
GN	CG11317.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
CC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OX	Ephyridiidae; Drosophilidae; Drosophila.			
RN	NCBI_TaxID=7227;			
RP	[1]			
RC	SEQUENCE FROM N.A.			
RX	STRAIN=BERKELEY;			
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
RA	Branden R.C., Rogers V., Blazek R.G., Champe M., Pfeiffer B.D.,			
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,			
RA	Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,			
RA	Balcer R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,			
RA	Beeoon K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,			
RA	Borkova D., Botchan M.R., Bouck J., Brinkstein P., Brotilier P.,			
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA	Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,			
RA	de Paulos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunov B.C., Dunn P.,			
RA	Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,			
RA	Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,			
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,			
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Jengam C.,			
RA	Jatelli M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,			
RA	Kamel B.E., Kodira C.D., Kraef C., Kravitz S., Kulp D., Lai Z.,			
RA	Lasko P.J., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,			
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,			
RA	Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,			

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spadling A.C., Stapleton M., Strong R., Sun E.,
 RA Styrkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "the genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003776; AAF57113.1; -
 DR FLYBASE: FBgn0039816; CG11317.
 DR INTERPRO: IPR000822; -
 DR PROSITE: PS00028; ZINC_FINGER_C2H2; 2.
 DR SEQUENCE 298 AA; 32241 MW; 58737FBF5B18C2E8 CRC64;

Query Match 100.0%; Score 33; DB 5; Length 298;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEANGN 6
 Db 95 GEANGN 100

RESULT 2
 ID Q27208 PRELIMINARY; PRT; 394 AA.
 AC Q27208;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
 DE IMMOBILIZATION ANTIGEN PRECURSOR (FRAGMENT).
 OS Ichthyophthirius multifiliis.
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
 OC Ophryoglenina; Ichthyophthirius.
 OX NCBI_TaxID=5932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GEORGIA;
 RX MEDLINE=92335298; PubMed=1631132;
 RA Clark T.G., McGraw R.A., Dickerson H.W.;
 RT "Developmental expression of surface antigen genes in the parasitic
 ciliate *Ichthyophthirius multifiliis*."
 RL Proc. Natl. Acad. Sci. U.S.A. 89:6363-6367(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GEORGIA;
 RX MEDLINE=93020590; PubMed=1383510;
 RA Lin T.L., Dickerson H.W.;
 RT "Purification and partial characterization of immobilization antigens
 from *Ichthyophthirius multifiliis*."
 RL J. Protozool. 39:457-463(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GEORGIA;
 RA Clark T.;
 RT Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GEORGIA;
 RA Clark T.;
 RT Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: M92907; AAC36158.1; -
 KW Signal.
 RL NON_TER 1 1 POTENTIAL.
 FT SIGNAL <1 1
 SQ SEQUENCE 394 AA; 39495 MW; 3013C2B2B8FDB682 CRC64;

Query Match 100.0%; Score 33; DB 5; Length 394;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEANGN 6
 Db 40 GEANGN 45

RESULT 3
 ID Q9NHC8 PRELIMINARY; PRT; 419 AA.
 AC Q9NHC8;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE LIM-HOMEODOMAIN TRANSCRIPTION FACTOR ISLET.
 OS Branchiostoma floridae (Florida lancelet) (Amphioxus).
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
 OC Branchiostoma.
 OX NCBI_TaxID=7739;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Jackman W.R., Langeland J.A., Kimmel C.B.;
 RT "Islet reveals segmentation in the amphioxus hindbrain homolog."
 RL Dev. Biol. 0:0-0(2000).
 DR EMBL: AF226616; AAF34717.1; -
 KW Homeobox; DNA-binding; Nuclear protein.
 SQ SEQUENCE 419 AA; 46140 MW; 0C11ID88A5BC1D88 CRC64;

Query Match 100.0%; Score 33; DB 5; Length 419;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEANGN 6
 Db 389 GEANGN 394

RESULT 4
 ID Q9XZG2 PRELIMINARY; PRT; 442 AA.
 AC Q9XZG2;
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
 DE IMMOBILIZATION ANTIGEN PRECURSOR.
 GN IAG48.
 OS Ichthyophthirius multifiliis.
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
 OC Ophryoglenina; Ichthyophthirius.
 OX NCBI_TaxID=5932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GI;
 RX MEDLINE=99196987; PubMed=10095108;
 RA Clark T.G., Lin T.L., Jackwood D.A., Sherrill J., Lin Y.,
 RA Dickerson H.W.;
 RT "The gene for an abundant parasite coat protein predicts tandemly
 repetitive metal binding domains."
 RL Gene 229:91-100(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GI;
 RA Gaertig J., Gao Y., Tishgarten T., Clark T.G., Dickerson H.W.;
 RT "Surface display of a parasite antigen in the ciliate *Tetrahymena*
thermophila."
 RL Nat. Biotechnol. 0:0-0(1999).
 DR EMBL: AF140273; AAD31283.1; -
 KW Signal.

FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 442 IMMOBILIZATION ANTIGEN.
SQ SEQUENCE 442 AA; 45025 MW; 52658F3F65D27AFA CRC64;

Query Match 100.0%; Score 33; DB 5; Length 442;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEANGN 6
Db 60 GEANGN 65

RESULT 5
O9PPH1 PRELIMINARY; PRT; 642 AA.
AC O9PPH1;
DT 01-OCT-2000 (TRENBLREL. 15, Created)
DT 01-OCT-2000 (TRENBLREL. 15, Last sequence update)
DT 01-OCT-2000 (TRENBLREL. 15, Last annotation update)
DE HYPOTHETICAL PROTEIN CJ0736.
GN CJ0736.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
CX Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC SFRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kelsey J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Felwell T., Holtroyd S.,
RA Jagers K., Karlyshev A., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., Van Vliet A.H.M.,
RA Whitehead S., Barrett B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
DR EMBL: AL139076; CAB73010.1; -.
KW Hypothetical protein.
SQ SEQUENCE 642 AA; 71612 MW; C047BECC71E9F44 CRC64;

Query Match 100.0%; Score 33; DB 2; Length 642;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEANGN 6
Db 197 GEANGN 202

RESULT 6
P73652 PRELIMINARY; PRT; 292 AA.
AC P73652;
DT 01-FEB-1997 (TRENBLREL. 02, Created)
DT 01-FEB-1997 (TRENBLREL. 02, Last sequence update)
DT 01-JUN-2000 (TRENBLREL. 14, Last annotation update)
DE HYPOTHETICAL 31.7 KDA PROTEIN.
GN SL11757.
OS Synechocystis sp. (strain PCC 6803)
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K., Okumura S.,
RA Shimpō S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;

RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions."
RL DNA Res. 3:109-136(1996).
DR EMBL: D90908; BAA17697.1; -.
KW Hypothetical protein.
SQ SEQUENCE 292 AA; 31717 MW; 7864DD212D9741BE CRC64;

Query Match 90.9%; Score 30; DB 2; Length 292;
Best Local Similarity 83.3%; Pred. No. 65;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEANGN 6
Db 184 GEANGN 189

RESULT 7
O92554 PRELIMINARY; PRT; 401 AA.
AC O92554;
DT 01-FEB-1997 (TRENBLREL. 02, Created)
DT 01-FEB-1997 (TRENBLREL. 02, Last sequence update)
DT 01-MAY-2000 (TRENBLREL. 13, Last annotation update)
DE MYELOBLAST KIAA0265 (FRAGMENT).
GN KIAA0265.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW;
RX MEDLINE=97191544; PubMed=9039502;
RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y., Ohara O.,
RA Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain."
RL DNA Res. 3:321-329(1996).
DR EMBL: D87454; BAA13395.1; -.
DR INTERPRO: IPR004798; -.
DR PFM: PF01344; Kelch; 2.
FT NON TER 1
SQ SEQUENCE 401 AA; 44378 MW; 81F753BE873F6D34 CRC64;

Query Match 90.9%; Score 30; DB 4; Length 401;
Best Local Similarity 83.3%; Pred. No. 93;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEANGN 6
Db 136 GEANGN 141

RESULT 8
O9X1X5 PRELIMINARY; PRT; 468 AA.
AC O9X1X5;
DT 01-NOV-1999 (TRENBLREL. 12, Created)
DT 01-NOV-1999 (TRENBLREL. 12, Last sequence update)
DT 01-OCT-2000 (TRENBLREL. 15, Last annotation update)
DE GLUTAMATE SYNTHASE, BETA SUBUNIT.
GN TM1640.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RX SFRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;

RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.,
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
genome sequence of *Thermotoga maritima*."
RL Nature 399:323-329(1999).
DR EMBL: AE001806; AAD36707.1; -.
DR HSSP: Q39243; 1VDC.
DR TIGR: TMA640; -.
DR INTERPRO: IPR000103; -.
DR INTERPRO: IPR000205; -.
DR INTERPRO: IPR000759; -.
DR INTERPRO: IPR001100; -.
DR INTERPRO: IPR001327; -.
DR PRINTS: PRO0368; PADPDR.
DR PRINTS: PRO0411; PNDPDTASEI.
DR PRINTS: PRO0419; ADXDPDTASE.
DR PRINTS: PRO0469; PNDPDTASEII.
SQ SEQUENCE 468 AA; 51613 MW; 062E26E1F28D07EA CRC64;

Query Match 90.9%; Score 30; DB 2; Length 468;
Best Local Similarity 83.3%; Pred. NO. 1.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEANGN 6
1:|||||
DB 345 GDANGN 350

RESULT 9
050396 PRELIMINARY; PRT; 588 AA.
AC 050396;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE PCRS-FAMILY PROTEIN.
GN RV3367 OR MTW004.25.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_Taxid=1773;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Kirogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of *Mycobacterium tuberculosis* from the
complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL: AL009198; CAI15752.1; -.
DR TUBERCULIST: RV3367; -.
DR INTERPRO: IPR000084; -.
DR INTERPRO: IPR002173; -.
DR PFAM: PF00934; PE: 1.
DR PROSITE: PSS0583; PFKB_KINASCS_1; UNKNOWN_1.
DR PRODOM: PD001223; -. 1.
SQ SEQUENCE 588 AA; 49708 MW; 067B84097F61DAF1 CRC64;

Query Match 90.9%; Score 30; DB 2; Length 588;
Best Local Similarity 83.3%; Pred. NO. 1.4e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 GEANGN 6
1:|||||
DB 287 GDANGN 292

RESULT 10
09VHC3 PRELIMINARY; PRT; 720 AA.
AC 09VHC3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CG3379 PROTEIN.
GN CG3379.
OS *Drosophila melanogaster* (fruit fly)
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Alameddine S., Baker S.E., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cacten E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleab J.M.,
RA Palazolo M., Pattan G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson W., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton W., Strong R., Sun E.,
RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveli J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*."
RL Science 287:2185-2195(2000).
DR EMBL: AE003683; AAF54393.1; -.
DR HSSP: P06241; 1AOU.
DR FLYBASE: FBgn0037704; CG9379.
DR INTERPRO: IPR000980; -.
DR PFAM: PF00017; SH2; 1.
DR PROSITE: PSS0001; SH2; 1.
SQ SEQUENCE 720 AA; 79468 MW; 7108BDF0E080BA CRC64;

Query Match 90.9%; Score 30; DB 5; Length 720;

Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEANGN 6
1:1111
DB 39 GGNNGN 44

RESULT 11
Q9N9R8 PRELIMINARY; PRT; 931 AA.

AC Q9N9R8; 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE HYPOTHETICAL 100.9 KDA PROTEIN.

OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN 11
RP SEQUENCE FROM N.A.

RA Tosato V., Bruschi C.V., Ciaroni L., Ivens A.C., Quail M.,
RA Rajadream M.A., Barrill B.G.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.
RC STRAIN=FRIDLIN;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).

DR EMBL: AL359217; CAB94664.1; -;
KW Hypothetical protein.
SQ SEQUENCE 931 AA; 100906 MW; 9C04C89B1180E2C4 CRC64;

Query Match 90.9%; Score 30; DB 5; Length 931;
Best Local Similarity 83.3%; Pred. No. 2.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEANGN 6
1:1111
DB 828 GGNNGN 833

RESULT 12
Q9SUX7 PRELIMINARY; PRT; 184 AA.

AC Q9SUX7; 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE HYPOTHETICAL 20.9 KDA PROTEIN.

GN F7K2.9 OR AT4G22420.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;
RN 11
RP SEQUENCE FROM N.A.
RA Bevan M., Medler H., Wambutt R., Bancroft I., Mewes H.W.,
RA Meyer K.F.X., Lemcke K., Schueller C.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

RN [3]
RP SEQUENCE FROM N.A.
RA Robben J., Grymonprez B., Volckaert G., Mewes H.W., Lemcke K.,

RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [4]
RP 1 SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL033845; CAB52816.1; -;
DR EMBL: AL161557; CAB79197.1; -;
KW Hypothetical protein.
SQ SEQUENCE 184 AA; 20867 MW; 8584E998D19C1C0D CRC64;

Query Match 87.9%; Score 29; DB 10; Length 184;
Best Local Similarity 83.3%; Pred. No. 64;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GEANGN 6
1:1111
DB 90 GGNNGN 95

RESULT 13
Q16398 PRELIMINARY; PRT; 217 AA.

AC Q16398; 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE C17E7.1 PROTEIN.

GN C17E7.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabdilita; Rhabditoidea;
OC Rhabdilitidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL NZ;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaison N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin R., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sprat J., Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

elegans.";
RL Nature 368:32-38(1994).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL NZ;
RA Birdsaw H.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL NZ;
RA Waterston R.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF016443; AAC24274.1; -;
DR INTERPRO: IPR001628; -;
DR PFAM: PF00105; zf-C4; 1.
DR PRINTS: PRO0047; STROIDFINGER.

SQ SEQUENCE 217 AA; 24724 MW; D239F24B38EF39D8 CRC64;

Query Match 87.9%; Score 29; DB 5; Length 217;
Best Local Similarity 83.3%; Pred. No. 77;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEANGN 6
1:1111
DB 90 GGNNGN 95

Db 19 GKANGN 24

RESULT 14

096716 PRELIMINARY; PRT; 391 AA.
 AC 096716;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE DOPAMINE D1/BETA RECEPTOR.
 OS Branchiostoma lanceolatum (Common lancelet) (Amphioxus).
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
 OC Branchiostoma.
 RN NCBL_TaxID=7740;
 RP SEQUENCE FROM N.A.
 RA Cardinaud B., Gilbert J.M., Sugamori K.S., Coudouel S., Gilbert B.,
 RA Vincent J.D., Niznik H.B., Vernier P.;
 RT "The amphioxus D1/beta receptor and the emergence of the vertebrate
 RT adrenergic system."
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ005433; CA06536.1; -
 DR INTERPRO: IPR000276; -
 DR INTERPRO: IPR000611; -
 DR INTERPRO: IPR000929; -
 DR INTERPRO: IPR000995; -
 DR INTERPRO: IPR002106; -
 DR INTERPRO: IPR002233; -
 DR PFAM: PF00001; 7tm_1; 1.
 DR PRINTS: PR00237; GPCRHHODPSN.
 DR PRINTS: PR00242; DOPAMINER.
 DR PRINTS: PR00243; MUSCARINICR.
 DR PRINTS: PR01012; NRPEPTIDEYR.
 DR PRINTS: PR01103; ADRENERGICR.
 DR PROSITE: PS00237; G-PROTEIN_RECEPTOR; UNKNOWN_1.
 DR PROSITE: PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
 DR Receptor.
 KW SEQUENCE 391 AA; 42622 MW; 67A5AD944AFA3FBE CRC64;

Query Match 87.9%; Score 29; DB 5; Length 391;
 Best Local Similarity 83.3%; Pred. No. 1.5e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEANGN 6
 1:||||
 Db 376 GKANGN 381

RESULT 15
 010052 PRELIMINARY; PRT; 473 AA.
 AC 010052;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE HYPOTHETICAL GLYCINE RECEPTOR LIKE PROTEIN F09C12.1 IN CHROMOSOME II.
 GN F09C12.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditicoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 RN NCBL_TaxID=6239;
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Anderson K.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL-MEMBRANE PROTEIN (PROBABLY).
 CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
 DR EMBL: U28929; AAA68348.1; -
 DR WORMPEP: F09C12.1; CE01901.

DR INTERPRO: IPR001175; -
 DR PFAM: PF00065; neur_chan; 1.
 DR PRINTS: PR00252; NRIONCHANNEL.
 DR PROSITE: PS00236; NEUROTR_ION_CHANNEL; UNKNOWN_1.
 KW Hypothetical protein; Receptor; Postsynaptic membrane; Ionic channel;
 KW Glycoprotein; Transmembrane.
 FT TRANSMEM 44 64
 FT TRANSMEM 287 307 POTENTIAL.
 FT TRANSMEM 311 331 POTENTIAL.
 FT TRANSMEM 350 370 POTENTIAL.
 FT DISULFID 195 195 BY SIMILARITY
 SO SEQUENCE 473 AA; 54322 MW; 128E13471C9FE8A9 CRC64;

Query Match 87.9%; Score 29; DB 5; Length 473;
 Best Local Similarity 83.3%; Pred. No. 1.9e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GEANGN 6
 1:||||
 Db 431 GETNGN 436

Search completed: March 6, 2001, 12:52:49
 Job time: 289 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:49:32 ; Search time 83.05 Seconds
(without alignments)
13.999 Million cell updates/sec

Title: US-09-196-161d-4

Perfect score: 176
Sequence: 1 PFANNAARAGICVPCQINRGSVTNAGDLATLAT 34

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

A_Geneseq_36.*
1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49.5	28.1	394	20 Y38623	Neisseria meningit
2	49.5	28.1	588	21 Y74776	Neisseria meningit
3	49.5	28.1	706	20 Y38624	Neisseria meningit
4	49	27.8	345	19 W59793	Neisseria meningit
5	49	27.8	353	19 W59790	Amino acid sequenc
6	49	27.8	354	19 W59791	Amino acid sequenc
7	49	27.8	354	20 Y41711	Human PRO719 prote
8	49	27.8	500	19 W59792	Amino acid sequenc
9	49	27.8	500	20 Y23759	Human endothelial
10	49	27.8	646	20 Y14942	Amino acid sequenc
11	49	27.8	646	20 Y40435	Human FATP protein
12	49	27.8	646	20 Y40436	Human FATP1 protei

13	49	27.8	646	20 Y14946	Amino acid sequenc
14	49	27.8	1157	20 W84589	Amino acid sequenc
15	48.5	27.6	706	21 Y74777	Neisseria meningit
16	48	27.3	1761	20 Y15457	Human laminin beta
17	47.5	27.0	3110	16 R71730	Merosin major subu
18	47.5	27.0	3110	20 Y15460	Human laminin alphi
19	47	26.7	161	16 R79553	Glutamic acid rele
20	46.5	26.4	243	17 W08094	Glutamic acid isol
21	46	26.1	637	20 Y25004	Melon MEERS protei
22	46	26.1	763	21 Y81501	Streptococcus pneu
23	46	26.1	764	21 Y81563	Streptococcus pneu
24	46	26.1	1477	14 R41724	High molecular wei
25	46	26.1	1477	14 R41728	High molecular wei
26	46	26.1	1477	15 R63506	Haemophilus high m
27	46	26.1	1477	18 W30294	Non-typable haemo
28	46	26.1	1785	20 Y15461	Human laminin beta
29	46	26.1	1786	19 W50893	Human laminin Bi c
30	45.5	25.9	400	20 Y38626	Neisseria gonorrhoe
31	45.5	25.9	400	21 Y74775	Neisseria gonorrhoe
32	45.5	25.9	615	17 W06830	Major neutralising
33	45.5	25.9	706	20 Y38627	Neisseria gonorrhoe
34	45	25.6	52	19 W59789	Neisseria gonorrhoe
35	45	25.6	75	19 W59794	Amino acid sequenc
36	45	25.6	405	20 Y14954	Amino acid sequenc
37	45	25.6	433	18 W14001	Enolase protein.
38	45	25.6	500	20 Y23760	Mouse endothelial
39	45	25.6	506	20 Y14934	Amino acid sequenc
40	45	25.6	643	20 Y14945	Amino acid sequenc
41	45	25.6	643	20 Y14958	Amino acid sequenc
42	45	25.6	652	18 W18010	Plasmodium falcipa
43	45	25.6	1156	19 W46857	Bacillus thuringie
44	45	25.6	1156	20 Y24960	Bacillus thuringie
45	45	25.6	1157	20 W84581	Amino acid sequenc

ALIGNMENTS

RESULT 1	
ID Y38623	standard; Protein: 394 AA.
XX Y38623;	
AC	
XX	
DT 08-OCT-1999	(first entry)
XX	
DE Neisseria meningitidis antigen encoded by a partial ORE64.	
XX	
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;	
XX treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.	
XX	
OS Neisseria meningitidis.	
XX	
PN W09924578-A2.	
XX	
PD 20-MAY-1999.	
XX	
PF 09-OCT-1998;	98WO-IB01665.
XX	
PR 01-SEP-1998;	98GB-0019016.
XX	
PR 06-NOV-1997;	97GB-0023516.
XX	
PR 14-NOV-1997;	97GB-0024190.
XX	
PR 18-NOV-1997;	97GB-0024386.
XX	
PR 27-NOV-1997;	97GB-0025158.
XX	
PR 10-DEC-1997;	97GB-0026147.
XX	
PR 14-JAN-1998;	98GB-0000759.
XX	
PA (CHIR-) CHIRON SPA.	
XX	
PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;	
XX WPT; 1999-327407/27.	
DR N-PSDB; Z12080.	

CC organisms are closely related. Fragments of the nucleic acids
 CC are useful as hybridisation probes and antisense reagents.
 XX
 SQ Sequence 706 AA;

Query Match 28.1%; Score 49.5; DB 20; Length 706;
 Best Local Similarity 44.4%; Pred. No. 52;
 Matches 12; Conservative 6; Mismatches 8; Indels 1; Gaps 1;

OY 3 ANNAARGICVPCQINRGSVTNAGDL 29
 11:11 111111:11:11:
 Db 128 aadnal-gnavpvqldl1gaas1pgdm 153

RESULT 4
 W59793
 ID W59793 standard; Protein; 345 AA.
 AC W59793;

DT 26-OCT-1998 (first entry)

DE Amino acid sequence of the catalytic domain of triglycerol lipase.

XX Human; triglycerol lipase; lipase like gene; LIG; hepatic lipase;
 KW human lipoprotein lipase; heparin; phosphatidylcholine ester;
 KM laundry detergent; serum lipid; atherosclerosis; diabetes;
 KW hyperlipidemia; intrahepatic cholestasis.

OS Homo sapiens.

XX

PN W09824888-A2.

PD 11-JUN-1998.

PF 05-DEC-1997; 97WO-US22331.

XX

PR 06-DEC-1996; 96US-0032783.

PR 06-DEC-1996; 96US-0032254.

XX

PA (RHON) RHONE-POULENC RORER PHARM INC.

XX

PI Amin DV, Doan KT, Jaye MC, Krawiec JA, Lynch KJ;

PI South VT;

XX

DR WPI: 1998-333310/29.

DR N-PSDB; V41623.

XX

PT Lipase like gene polypeptides - used for hydrolysis of

PT phosphatidylcholine esters or for treating e.g. atherosclerosis,

PT diabetes, hyperlipidemia or intrahepatic cholestasis

XX

PS Claim 2; Pages 60-61; 94pp; English.

XX

CC This is the amino acid sequence of the catalytic domain of the

CC triglycerol lipase used in the method of the invention involving the

CC human lipase like gene (LIG). LIG has homology with human lipoprotein

CC lipase and hepatic lipase, and binds to heparin. The LIG polypeptides

CC can be used for the enzymatic hydrolysis of phosphatidylcholine esters,

CC for e.g. industrial or food processing, or in laundry detergents. The

CC products can also be used for improving the serum lipid profiles of

CC animals, e.g. in the treatment of atherosclerosis, diabetes,

CC hyperlipidemia or intrahepatic cholestasis. The products can also be

CC used for detection, diagnosis and drug screening.

XX

SQ Sequence 345 AA;

Query Match 27.8%; Score 49; DB 19; Length 345;

Best Local Similarity 40.0%; Pred. No. 28;

Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 4 ANNAARGICVPCQINRGSV 23
 11:111111:1111:
 Db 300 snrfkkgjclscrkncnsi 319

RESULT 5
 W59790
 ID W59790 standard; Protein; 353 AA.
 AC W59790;

DT 26-OCT-1998 (first entry)

DE Amino acid sequence of lipase like protein (LIG).

XX Human lipase like gene; LIG; human lipoprotein lipase; hepatic lipase;
 KW heparin; phosphatidylcholine ester; laundry detergent; serum lipid;
 KM atherosclerosis; diabetes; hyperlipidemia; intrahepatic cholestasis.

OS Homo sapiens.

XX

PN W09824888-A2.

PD 11-JUN-1998.

PF 05-DEC-1997; 97WO-US22331.

XX

PR 06-DEC-1996; 96US-0032783.

PR 06-DEC-1996; 96US-0032254.

XX

PA (RHON) RHONE-POULENC RORER PHARM INC.

XX

PI Amin DV, Doan KT, Jaye MC, Krawiec JA, Lynch KJ;

PI South VT;

XX

DR WPI: 1998-333310/29.

DR N-PSDB; V41620.

XX

PT Lipase like gene polypeptides - used for hydrolysis of

PT phosphatidylcholine esters or for treating e.g. atherosclerosis,

PT diabetes, hyperlipidemia or intrahepatic cholestasis

XX

PS Disclosure; Pages 47-49; 94pp; English.

XX

CC This is the amino acid sequence of the human lipase like protein (LIG).

CC It has homology with human lipoprotein lipase and hepatic lipase, and

CC binds to heparin. The LIG polypeptides can be used for the enzymatic

CC hydrolysis of phosphatidylcholine esters, for e.g. industrial or food

CC processing, or in laundry detergents. The products can also be used

CC for improving the serum lipid profiles of animals, e.g. in the

CC treatment of atherosclerosis, diabetes, hyperlipidemia or

CC intrahepatic cholestasis. The products can also be used for detection,

CC diagnosis and drug screening.

XX

SQ Sequence 353 AA;

Query Match 27.8%; Score 49; DB 19; Length 353;

Best Local Similarity 40.0%; Pred. No. 29;

Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 4 ANNAARGICVPCQINRGSV 23
 11:111111:1111:
 Db 300 snrfkkgjclscrkncnsi 319

RESULT 6
 W59791
 ID W59791 standard; Protein; 354 AA.
 AC W59791;

DT 26-OCT-1998 (first entry)

DE XX Amino acid sequence of lipase like protein LIGN.
XX
KW Human lipase like gene; LIGN; human lipoprotein lipase; hepatic lipase;
KW heparin; phosphatidylcholine ester; laundry detergent; serum lipid;
XX atherosclerosis; diabetes; hyperlipidemia; intrahepatic cholestasis.
OS Homo sapiens.
XX
PN WO9824888-A2.
XX
PD 11-JUN-1998.
XX
PE 05-DEC-1997; 97WO-US22331.
XX
PR 06-DEC-1996; 96US-0032783.
PR 06-DEC-1996; 96US-0032254.
XX
PA (RHON) RHONE-POULENC RORER PHARM INC.
XX
PI Amin DV, Doan KT, Jaye MC, Krawiec JA, Lynch KJ;
PI South VT;
XX
DR WPI: 1998-333310/29.
DR N-PSDB: V41621.
XX
PT Lipase like gene polypeptides - used for hydrolysis of
PT phosphatidylcholine esters or for treating e.g. atherosclerosis,
PT diabetes, hyperlipidemia or intrahepatic cholestasis
XX
PS Claim 6; Pages 21-52; 94pp; English.
XX
CC This is the amino acid sequence of the human lipase like protein (LIGN).
CC It has homology with human lipoprotein lipase and hepatic lipase, and
CC binds to heparin. The LIGN polypeptides can be used for the enzymatic
CC hydrolysis of phosphatidylcholine esters, for e.g. industrial or food
CC processing, or in laundry detergents. The products can also be used
CC for improving the serum lipid profiles of animals, e.g. in the
CC treatment of atherosclerosis, diabetes, hyperlipidemia or
CC intrahepatic cholestasis. The products can also be used for detection,
CC diagnosis and drug screening.
XX
SQ Sequence 354 AA;

Query Match 27.8%; Score 49; DB 19; Length 354;
Best Local Similarity 40.0%; Pred. No. 29;
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 4 ANNAAGICVPCQINRVGSV 23
: :|||: :| :
Db 300 snrfkkgjlcscrkrcnsi 319

RESULT 7
Y41711
ID Y41711 standard; Protein; 354 AA.
XX
XX Y41711;
AC
XX
DT 07-DEC-1999 (first entry)
XX
DE Human PRO719 protein sequence.
XX
KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW secreted protein; transmembrane protein.
XX
XX Homo sapiens.
XX
XX OS
XX PN WO9946281-A2.
XX
XX PD 16-SEP-1999.

XX
PF 08-MAR-1999; 99WO-US05028.
XX
PR 10-MAR-1998; 98US-0077450.
PR 11-MAR-1998; 98US-0077632.
PR 11-MAR-1998; 98US-0077641.
PR 11-MAR-1998; 98US-0077649.
PR 12-MAR-1998; 98US-0077791.
PR 13-MAR-1998; 98US-0078004.
PR 17-MAR-1998; 98US-0040220.
PR 20-MAR-1998; 98US-0078886.
PR 20-MAR-1998; 98US-0078910.
PR 20-MAR-1998; 98US-0078936.
PR 20-MAR-1998; 98US-0078939.
PR 25-MAR-1998; 98US-0079294.
PR 26-MAR-1998; 98US-0079656.
PR 27-MAR-1998; 98US-0079653.
PR 27-MAR-1998; 98US-0079664.
PR 27-MAR-1998; 98US-0079689.
PR 27-MAR-1998; 98US-0079728.
PR 27-MAR-1998; 98US-0079786.
PR 30-MAR-1998; 98US-0079920.
PR 30-MAR-1998; 98US-0079923.
PR 31-MAR-1998; 98US-0080105.
PR 31-MAR-1998; 98US-0080107.
PR 31-MAR-1998; 98US-0080165.
PR 31-MAR-1998; 98US-0080194.
PR 01-APR-1998; 98US-0080327.
PR 01-APR-1998; 98US-0080328.
PR 01-APR-1998; 98US-0080333.
PR 01-APR-1998; 98US-0080334.
PR 08-APR-1998; 98US-0081049.
PR 08-APR-1998; 98US-0081070.
PR 08-APR-1998; 98US-0081071.
PR 09-APR-1998; 98US-0081195.
PR 09-APR-1998; 98US-0081203.
PR 09-APR-1998; 98US-0081229.
PR 15-APR-1998; 98US-0081817.
PR 15-APR-1998; 98US-0081838.
PR 15-APR-1998; 98US-0081952.
PR 15-APR-1998; 98US-0081955.
PR 21-APR-1998; 98US-0082568.
PR 21-APR-1998; 98US-0082569.
PR 22-APR-1998; 98US-0082700.
PR 22-APR-1998; 98US-0082704.
PR 22-APR-1998; 98US-0082804.
PR 23-APR-1998; 98US-0082767.
PR 23-APR-1998; 98US-0082796.
PR 27-APR-1998; 98US-0083336.
PR 28-APR-1998; 98US-0083392.
PR 29-APR-1998; 98US-0083392.
PR 29-APR-1998; 98US-0083405.
PR 29-APR-1998; 98US-0083496.
PR 29-APR-1998; 98US-0083499.
PR 29-APR-1998; 98US-0083500.
PR 29-APR-1998; 98US-0083545.
PR 29-APR-1998; 98US-0083554.
PR 29-APR-1998; 98US-0083558.
PR 30-APR-1998; 98US-0083559.
PR 30-APR-1998; 98US-0083742.
PR 05-MAY-1998; 98US-0084366.
PR 06-MAY-1998; 98US-0084414.
PR 06-MAY-1998; 98US-0084414.
PR 07-MAY-1998; 98US-0084598.
PR 07-MAY-1998; 98US-0084600.
PR 07-MAY-1998; 98US-0084627.
PR 07-MAY-1998; 98US-0084637.
PR 07-MAY-1998; 98US-0084639.
PR 07-MAY-1998; 98US-0084640.
PR 07-MAY-1998; 98US-0084643.
PR 13-MAY-1998; 98US-0085323.
PR 13-MAY-1998; 98US-0085338.
PR 13-MAY-1998; 98US-0085339.

PR 15-MAY-1998; 98US-0085573.
 PR 15-MAY-1998; 98US-0085579.
 PR 15-MAY-1998; 98US-0085580.
 PR 15-MAY-1998; 98US-0085582.
 PR 15-MAY-1998; 98US-0085589.
 PR 15-MAY-1998; 98US-0085697.
 PR 15-MAY-1998; 98US-0085700.
 PR 15-MAY-1998; 98US-0085704.
 PR 18-MAY-1998; 98US-0086023.
 PR 22-MAY-1998; 98US-0086392.
 PR 22-MAY-1998; 98US-0086414.
 PR 22-MAY-1998; 98US-0086430.
 PR 28-MAY-1998; 98US-0086486.
 PR 28-MAY-1998; 98US-0087098.
 PR 28-MAY-1998; 98US-0087106.
 PR 28-MAY-1998; 98US-0087208.
 PR 30-JUL-1998; 98US-0094651.
 PR 11-SEP-1998; 98US-0100038.

XX (GETH) GENENTECH INC.

PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;

XX WPI: 1999-551358/46.

DR N-PSDB; 234041.

XX New secreted and transmembrane polypeptides and their polynucleotides,
 PT useful for treating blood coagulation disorders, cancers and cellular
 PT adhesion disorders.

XX Claim 12; Fig 66; 530pp; English.

XX The present invention describes secreted and transmembrane polypeptides
 CC and their polynucleotides. The nucleotide sequences are useful as
 CC sources of probes, primers, for chromosome mapping, and for generation
 CC of antisense sequences. They can also be used to create transgenic
 CC animals. The proteins can be used to treat a variety of diseases and
 CC disorders, depending on their function. Diseases that may be treated
 CC include blood coagulation disorders, cancers and cellular adhesion
 CC disorders. They may also be used to raise antibodies. 233891 to
 CC 234338, and 241685 to 241774 represent polynucleotide and polypeptide
 CC sequence given in the exemplification of the present invention.

XX Sequence 354 AA;

Query Match 27.8%; Score 49; DB 20; Length 354;

Best Local Similarity 40.0%; Pred. NO. 29;

Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 4 ANNMARGICVPCOINRGVS 23

DB 300 snrfkkgiclsrknrcnsi 319

RESULT 8

W59792 W59792 standard; Protein; 500 AA.

XX W59792;

XX 26-OCT-1998 (first entry)

XX Amino acid sequence of lipase like protein LIGXL.

XX Human lipase like gene; LIGXL; human lipoprotein lipase; hepatic lipase;
 KW heparin; phosphatidylcholine ester; laundry detergent; serum lipid;
 KW atherosclerosis; diabetes; hyperlipidemia; intrahepatic cholestasis.

XX Homo sapiens.

XX W09824888-A2.

PD 11-JUN-1998.

XX 05-DEC-1997; 97WO-US22331.

XX 06-DEC-1996; 96US-0032783.

PR 06-DEC-1996; 96US-0032254.

XX (RHON) RHONE-POULENC RORER PHARM INC.

PI Amin DV, Doan KT, Jaye MC, Krawiec JA, Lynch KI;

PI South VJ;

XX WPI: 1998-333310/29.

DR N-PSDB; VA1692.

PT Lipase like gene polypeptides - used for hydrolysis of
 PT phosphatidylcholine esters or for treating e.g. atherosclerosis,
 PT diabetes, hyperlipidemia or intrahepatic cholestasis

XX Claim 3; Pages 56-58; 94pp; English.

XX This is the amino acid sequence of the human lipase like protein (LIGXL).
 CC It has homology with human lipoprotein lipase and hepatic lipase, and
 CC binds to heparin. The LIGXL polypeptides can be used for the enzymatic
 CC hydrolysis of phosphatidylcholine esters, for e.g. industrial or food
 CC processing, or in laundry detergents. The products can also be used
 CC for improving the serum lipid profiles of animals e.g. in the
 CC treatment of atherosclerosis, diabetes, hyperlipidemia or
 CC intrahepatic cholestasis. The products can also be used for detection,
 CC diagnosis and drug screening.

XX Sequence 500 AA;

Query Match 27.8%; Score 49; DB 19; Length 500;

Best Local Similarity 40.0%; Pred. NO. 42;

Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 4 ANNMARGICVPCOINRGVS 23

DB 300 snrfkkgiclsrknrcnsi 319

RESULT 9

Y23759 Y23759 standard; Protein; 500 AA.

XX Y23759;

XX 09-SEP-1999 (first entry)

XX Human endothelial cell lipase protein sequence.

XX Endothelial lipase; hypertriglyceridemia; hypercholesterolemia;
 KW cancer; de nova imaging; endothelial cell; tumour; vascular disorder;
 KW lipidemia; diabetes; obesity; restenosis; cancer.

XX Homo sapiens.

XX W09932611-A1.

XX 01-JUL-1999.

XX 21-DEC-1998; 98WO-US27335.

XX 19-DEC-1997; 97US-0068336.

XX (PROG-) PROGENITOR INC.

XX (UYVA-) UNIV VANDERBILT.

XX Cioffi JA, Hirata K, Quartermous T, Zupancic T;

DR WPI: 1999-418920/35.

CC intracellular or blood levels of long chain fatty acids. Such compounds
CC are especially useful to treat conditions associated with deficient
CC regulation (e.g. may comprise an inhibitor to treat cardiomyopathies or
CC diabetes or an enhancer to treat obesity. The polynucleotides are also
CC useful to screen compounds for their effects on hFATP expression, e.g.
CC by measuring mRNA transcription in cells/cell extracts (e.g. liver
CC cells) contacted with the compound and comparing with that in non-
CC contacted cells. The present sequence represents the hFATP protein.

XX Sequence 646 AA;

Query Match 27.8%; Score 49; DB 20; Length 646;
Best Local Similarity 50.0%; Pred. No. 56;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 8 ARGICVPCQINRVG 21
|:|:|:|:|:|:|
Db 435 agglcipcagepg 448

RESULT 12

ID Y40436 standard; Protein; 646 AA.

XX Y40436;

AC Y40436;

XX 08-FEB-2000 (first entry)

DE Human FATP1 protein sequence.

XX Fatty acid transport protein; FATP; hFATP1; cardiomyopathy; diabetes;

KW long-chain fatty acid metabolism; obesity; human.

XX Homo sapiens.

OS W09951740-A2.

XX 14-OCT-1999.

XX 02-APR-1999; 99WO-EP02295.

XX 06-APR-1998; 98EP-0400823.

XX (JANSEN PHARM NV.

XX (UNIM) UNIV WASHINGTON.

XX Martin G, Nemoto M, Deeb SS, Auwerx J;

XX WPI: 1999-620202/53.

XX New human fatty acid transport protein, hFATP, useful to screen for

XX inhibitors or enhancers useful to regulate fatty acid metabolism

XX Claim 1; Fig 2; 83pp; English.

XX The invention provides a human fatty acid transport protein (hFATP).

XX hFATP is believed to be involved in the modulation long-chain fatty acid

XX metabolism; the protein and polynucleotides therefore enable production

XX of compositions comprising a component regulating (inhibiting or

XX enhancing) expression of the hFATP gene useful therapeutically to alter

XX intracellular or blood levels of long chain fatty acids. Such compounds

XX are especially useful to treat conditions associated with deficient

XX regulation (e.g. may comprise an inhibitor to treat cardiomyopathies or

XX diabetes or an enhancer to treat obesity. The polynucleotides are also

XX useful to screen compounds for their effects on hFATP expression, e.g.

XX by measuring mRNA transcription in cells/cell extracts (e.g. liver

XX cells) contacted with the compound and comparing with that in non-

XX contacted cells. The present sequence represents the hFATP1 protein.

XX Sequence 646 AA;

Query Match 27.8%; Score 49; DB 20; Length 646;
Best Local Similarity 50.0%; Pred. No. 56;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 8 ARGICVPCQINRVG 21
|:|:|:|:|:|:|
Db 435 agglcipcagepg 448

RESULT 13

ID Y14946 standard; protein; 646 AA.

XX Y14946;

AC Y14946;

XX 26-OCT-1999 (first entry)

DE Amino acid sequence of human hFATP1.

XX Fatty acid transport protein; FATP; long chain fatty acid; LCFA; human;

KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease.

XX Homo sapiens.

OS W09936537-A2.

XX 22-JUL-1999.

XX 14-JAN-1999; 99WO-US00182.

XX 14-JAN-1999; 99US-0232201.

XX 15-JAN-1998; 98US-0071374.

XX 20-JUL-1998; 98US-0093491.

XX 04-DEC-1998; 98US-0110941.

XX 14-JAN-1999; 99US-0232195.

XX 14-JAN-1999; 99US-0232197.

XX 14-JAN-1999; 99US-0232200.

XX (MILL-) MILLENNIUM PHARM INC.

XX (WHED) WHITEHEAD INST BIOMEDICAL RES.

XX Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;

XX WPI: 1999-444398/37.

XX N-PSDB; 200356.

XX Fatty acid transport proteins and related polynucleotides, useful

XX for treating obesity, diabetes and heart disease

XX Claim 30; Fig 45; 255pp; English.

XX The invention provides a family of fatty acid transport proteins (FATPs)

XX that mediate transport of long chain fatty acids (LCFAs) across cell

XX membranes into cells. Human and murine FATP proteins and nucleic acids

XX encoding the proteins are provided. The FATP proteins can be produced

XX by standard recombinant methodology. Fatty acid uptake by cells can be

XX modulated by modulating biosynthesis of FATP proteins especially FATP6.

XX In particular, antisense oligonucleotides can be used to modulate FATP

XX biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid

XX uptake in cardiac muscle of humans. Agents can be directed to cardiac

XX muscle or liver by administration of a complex of the agent and a FATP6

XX binding moiety. DNA encoding FATP proteins can be used as a reference

XX used in detecting variant alleles or homologues. Altering the LCFA uptake

XX by administering an inhibitor or enhancer of FATP transport function in

XX the small intestine can decrease or increase calories available as fats,

XX and can decrease or increase circulating fatty acids. Blocking the

XX function of FATP4 and also FATP2, is useful for treating obesity,

XX diabetes and heart disease.

XX Sequence 646 AA;

Query Match 27.8%; Score 49; DB 20; Length 646;

Best Local Similarity 50.0%; Pred. No. 56;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
OY 8 ARGICVPCQINRVG 21
1:1:1:1:1:1
Db 435 agjlcipcagepg 448

RESULT 14
W84589
ID W84589 standard; Protein; 1157 AA.

XX W84589;

DT 07-JUN-1999 (first entry)

DE Amino acid sequence of *Bacillus thuringiensis* Cry9C mutant 23.

XX Cry9C; crystal protein; toxic; mutant; insecticide; Insect;

KW transgenic plant; transgenic seed; Lepidoptera.

XX *Bacillus thuringiensis*.

OS Synthetic.

XX W09900407-A2.

PD 07-JAN-1999.

PF 25-JUN-1998; 98WO-EP04033.

PR 27-JUN-1997; 97US-0884389.

PA (PLBZ) PLANT GENETIC SYSTEMS NV.

PI De Roeck S, Van Rle J;

DR WPI; 1999-105666/09.

XX New *Bacillus thuringiensis* Cry9C crystal proteins with improved
PT toxicity - useful for producing transgenic plants in controlling
PT pest insects, especially *Ostrinia nubilalis*, *Heliothis virescens* and
PT *Diatraea grandiosella*

XX Examples ; Page -; 40pp; English.

XX This is the amino acid sequence of a Cry9C (crystal) mutant protein
CC which is toxic to an insect species. It is modified in the method of
CC the invention in an attempt to improve its toxicity. The improved
CC proteins are useful for producing transgenic plants and seeds in
CC controlling insects feeding on a plant. The transgenic plants are
CC also useful for producing more transgenic plants with similar
CC characteristics, or for introducing the improved cry9C gene
CC (including fragments) into the same plant variety or related plant
CC species. The Cry9C protein can be obtained from cultured cells of
CC the transgenic plants and used in insecticide compositions against
CC lepidopteran insects, especially *Ostrinia nubilalis*, *Heliothis*
CC *virescens* and *Diatraea grandiosella*. The cry9C gene can be
CC expressed in combination with another insect control protein (e.g.
CC Another *B. thuringiensis* derived crystal protein, especially CryIAb-
CC or CryIb-type protein), and used to transform *B. thuringiensis*, which
CC produces other insecticidal toxins. Such strains are useful against
CC a variety of insect pests, or insects where insect resistance
CC development is prevented or delayed.
CC (Note: this sequence is not given in the specification, but is
CC generated using the information provided by the inventors.)

XX Sequence 1157 AA;

Query Match 27.8%; Score 49; DB 20; Length 1157;
Best Local Similarity 56.2%; Pred. No. 1.1e+02;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 16 QINRVGSVTNAGDLAT 31
1:1:1:1:1:1
Db 481 qngvgyslanagsvpt 496

RESULT 15
Y74777
ID Y74777 standard; Protein; 706 AA.

XX Y74777;

DT 21-MAR-2000 (first entry)

DE *Neisseria meningitidis* ORF 258 protein sequence SEQ ID NO:1028.

XX *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;

KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;

XX antibacterial; gene therapy.

OS *Neisseria meningitidis*.

XX W09957280-A2.

PD 11-NOV-1999.

PF 30-APR-1999; 99WO-US09346.

PR 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998; 98US-0098994.

PR 02-SEP-1998; 98US-0099062.

PR 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.

PR 25-FEB-1999; 99US-0121528.

XX (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;

PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;

PI Tettelein H, Venter JC;

XX WPI; 2000-062150/05.

DR N-PSDB; 253539.

XX Novel *Neisseria* polypeptides predicted to be useful antigens for

PT vaccines and diagnostics

XX Claim 2; Page 597; 1453pp; English.

XX 253015 to 254536, 254577 to 254615, and Y74253 to Y75941 represent

CC novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides and

CC polypeptides. 254537 to 254576 and 254616 to 255473 represent PCR

CC primers used in the exemplification of the present invention. The

CC polypeptides, the polynucleotides, antibodies and compositions of

CC the invention can be used as vaccines, as diagnostic reagents, and as

CC immunogenic compositions. The polypeptides can be used in the

CC manufacture of medicaments for treating or preventing infection due to

CC *Neisseria* bacteria (e.g. meningitis and septicemia), to detect the

CC presence of *Neisseria* bacteria, or to raise antibodies. They may also

CC be used to screen for agonists or antagonists, which may themselves

CC have use as antibacterial agents. The polynucleotides of the invention

CC may also be used in gene therapy protocols.

XX Sequence 706 AA;

Query Match 27.6%; Score 48.5; DB 21; Length 706;
Best Local Similarity 40.7%; Pred. No. 72;
Matches 11; Conservative 7; Mismatches 8; Indels 1; Gaps 1;

OY 3 AANNARGICVPCQINRVGSVTNAGDL 29

• Fri-Mar 9 15:47:49 2001

us-09-196-161d-4.rag

Page 9

||: || | : ||: ||: ||:
Db 128 aadnal-gna1pvg1d1lgas1pgdm 153

Search completed: March 6, 2001, 12:49:33
Job time: 94 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using SW model

Run on: March 6, 2001, 12:53:47 ; Search time 57.76 Seconds
(Without alignments)
10.570 Million cell updates/sec

Title: US-09-196-161d-4
Perfect score: 176
Sequence: 1 PFANNMARGICVPCQINRVGSVTNAGDLATLAT 34

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/PCRTUS.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	30.1	314	2	US-08-460-309-19 Sequence 19, Appl
2	53	30.1	314	2	US-08-125-077-19 Sequence 19, Appl
3	48	27.3	846	2	US-07-728-215-33 Sequence 33, Appl
4	47.5	27.0	3111	2	US-08-460-309-4 Sequence 4, Appl
5	47.5	27.0	3111	2	US-08-125-077-4 Sequence 4, Appl
6	47	26.7	273	1	US-08-152-019A-30 Sequence 30, Appl
7	46	26.1	278	2	US-08-460-309-13 Sequence 13, Appl
8	46	26.1	278	2	US-08-125-077-13 Sequence 13, Appl
9	46	26.1	279	1	US-08-152-019A-29 Sequence 29, Appl
10	46	26.1	1196	1	US-08-144-121-4 Sequence 4, Appl
11	46	26.1	1196	2	US-08-735-893-4 Sequence 4, Appl
12	46	26.1	1477	1	US-08-038-682-4 Sequence 4, Appl
13	46	26.1	1477	1	US-08-302-832-4 Sequence 4, Appl
14	46	26.1	1477	2	US-08-530-198-4 Sequence 4, Appl
15	46	26.1	1477	2	US-08-469-880-4 Sequence 4, Appl
16	46	26.1	1477	2	US-08-728-470-4 Sequence 4, Appl
17	46	26.1	1477	2	US-08-617-697-4 Sequence 4, Appl
18	45.5	25.9	615	2	US-08-663-566A-17 Sequence 17, Appl
19	45.5	25.9	615	2	US-08-023-610-17 Sequence 17, Appl
20	45.5	25.9	615	2	US-08-288-065A-17 Sequence 17, Appl
21	45.5	25.9	615	2	US-08-362-240A-17 Sequence 17, Appl
22	45.5	25.9	615	4	PCT-US95-10245-17 Sequence 17, Appl
23	45	25.6	269	2	US-08-460-309-14 Sequence 14, Appl
24	45	25.6	269	2	US-08-125-077-14 Sequence 14, Appl
25	45	25.6	271	1	US-08-152-019A-28 Sequence 28, Appl
26	45	25.6	625	2	US-08-532-547-7 Sequence 7, Appl
27	45	25.6	625	2	US-08-532-547-9 Sequence 9, Appl
28	45	25.6	625	3	US-09-019-809-7 Sequence 7, Appl

29	45	25.6	625	3	US-09-019-809-9	Sequence 9, Appl
30	45	25.6	1157	2	US-08-532-547-5	Sequence 5, Appl
31	45	25.6	1157	2	US-08-379-656B-5	Sequence 5, Appl
32	45	25.6	1157	3	US-08-455-838-5	Sequence 5, Appl
33	45	25.6	1157	3	US-09-019-809-5	Sequence 5, Appl
34	43.5	24.7	313	1	US-08-302-449-2	Sequence 2, Appl
35	43.5	24.7	313	4	PCT-US94-07430-2	Sequence 2, Appl
36	43	24.4	120	2	US-08-232-087A-8	Sequence 8, Appl
37	43	24.4	156	3	US-08-600-982-30	Sequence 30, Appl
38	43	24.4	156	4	PCT-US94-10261A-30	Sequence 30, Appl
39	43	24.4	595	2	US-08-232-087A-2	Sequence 2, Appl
40	43	24.4	1248	2	US-08-348-953-17	Sequence 17, Appl
41	43	24.4	1248	2	US-08-465-965-17	Sequence 17, Appl
42	43	24.4	1248	3	US-08-465-966-17	Sequence 17, Appl
43	43	24.4	1713	3	US-08-600-982-24	Sequence 24, Appl
44	43	24.4	1713	4	PCT-US94-10261A-24	Sequence 24, Appl
45	42	23.9	169	2	US-08-460-309-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1
US-08-460-309-19
Sequence 19, Application US/08460309
Patent No. 5837496
GENERAL INFORMATION:
APPLICANT: Engvall, Eva
APPLICANT: Leivo, Ilmo
TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
NUMBER OF INVENTION: Fragments and Uses Thereof
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,309
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/125,077
FILING DATE: 22-SEP-1993
APPLICATION NUMBER: US PCT/US 94/10730
FILING DATE: 21-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/472,319
FILING DATE: 30-JAN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/919,951
FILING DATE: 27-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 9721
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 314 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-460-309-19

FILED DATE: 435
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/125,077
FILING DATE: 22-SEP-1993
APPLICATION NUMBER: US PCT/US 94/10730
FILING DATE: 21-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/472,319
FILING DATE: 30-JAN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/919,951
FILING DATE: 27-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 9721
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3111 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-460-309-4

Query Match 27.0%; Score 47.5; DB 2; Length 3111;
Best Local Similarity 62.5%; Pred. No. 4.2e+02;
Matches 10; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

OY 7 AARGICVPCQINRVS 22
: ||| ||| | | |
DB 960 SARG-CVPCNCSFGS 974

RESULT 5
US-08-125-077-4
Sequence 4, Application US/08125077
Patent No. 5872231
Patent No. 5872231 5840863
GENERAL INFORMATION:
APPLICANT: Engvall, Eva
APPLICANT: Leivo, Ilmo
TITLE OF INVENTION: Nucleic Acids Encoding Metrosin, Metrosin
TITLE OF INVENTION: Fragments and Uses Thereof
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/125,077
FILING DATE: 22-SEP-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US 94/10730
FILING DATE: 21-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/472,319
FILING DATE: 30-JAN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/919,951
FILING DATE: 27-JUL-1992

ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 9721
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3111 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-125-077-4

Query Match 27.0%; Score 47.5; DB 2; Length 3111;
Best Local Similarity 62.5%; Pred. No. 4.2e+02;
Matches 10; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

OY 7 AARGICVPCQINRVS 22
: ||| ||| | | |
DB 960 SARG-CVPCNCSFGS 974

RESULT 6
US-08-152-019A-30
Sequence 30, Application US/08152019A
Patent No. 5565331
GENERAL INFORMATION:
APPLICANT: Tessier-Lavigne, Marc
APPLICANT: Serafini, Tito
APPLICANT: Kennedy, Timothy
APPLICANT: Placzek, Marysia
APPLICANT: Jessell, Thomas
APPLICANT: Dodd, Jane
TITLE OF INVENTION: NEURAL AXON OUTGROWTH MODULATORS
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/152,019A
FILING DATE: 12-NOV-1993
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard Aron
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59012/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1969
TELEFAX: (415) 398-3249
TELEX: 910 277299 FHT UR
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 273 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-152-019A-30

Query Match 26.7%; Score 47; DB 1; Length 273;
Best Local Similarity 43.8%; Pred. No. 36;

COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/152,019A
FILING DATE: 12-NOV-1993
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard Aron
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59012/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299 FHT UR
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 279 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-152-019A-29

Query Match 26.1%; Score 46; DB 1; Length 279;
Best Local Similarity 42.1%; Pred. No. 50;
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 3 AANNAARGICVPCOINRVG 21
DB 84 ATGNVSGVCDDCOHNTMG 102

RESULT 10
US-08-144-121-4
Sequence 4, Application US/08144121
Patent No. 5610031
GENERAL INFORMATION:
APPLICANT: Burgeson, Robert E.
APPLICANT: Magman, David W.
TITLE OF INVENTION: BLK CHAIN OF LAMININ AND METHODS OF USE
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: BOSTON
STATE: Massachusetts
COUNTRY: United States
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/144,121
FILING DATE: 27-OCT-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Paul L.
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: (MGH-0780.0) MGP-021
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1196 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
FEATURE:
NAME/KEY: Domain
LOCATION: 1..250
FEATURE:
NAME/KEY: Domain
LOCATION: 251..437
FEATURE:
NAME/KEY: Domain
LOCATION: 438..807
FEATURE:
NAME/KEY: Domain
LOCATION: 808..840
FEATURE:
NAME/KEY: Domain
LOCATION: 841..1196
US-08-144-121-4

Query Match 26.1%; Score 46; DB 1; Length 1196;
Best Local Similarity 42.1%; Pred. No. 2,4e+02;
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 3 AANNAARGICVPCOINRVG 21
DB 333 ATGNVSGVCDDCOHNTMG 351

RESULT 11
US-08-735-893-4
Sequence 4, Application US/08735893
Patent No. 5914317
GENERAL INFORMATION:
APPLICANT: Burgeson, Robert E.
APPLICANT: Magman, David W.
TITLE OF INVENTION: BLK CHAIN OF LAMININ AND METHODS OF USE
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: BOSTON
STATE: Massachusetts
COUNTRY: United States
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/735,893
FILING DATE: 18-OCT-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/144,121
FILING DATE: 27-OCT-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Paul L.
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: (MGH-0780.1) MGP-021DV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1196 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide

FRAGMENT TYPE: internal
FEATURE:
NAME/KEY: Domain
LOCATION: 1..250
FEATURE:
NAME/KEY: Domain
LOCATION: 251..437
FEATURE:
NAME/KEY: Domain
LOCATION: 438..807
FEATURE:
NAME/KEY: Domain
LOCATION: 808..840
FEATURE:
NAME/KEY: Domain
LOCATION: 841..1196
US-08-735-893-4

Query Match 26.1%; Score 46; DB 2; Length 1196;
Best Local Similarity 42.1%; Pred. No. 2.4e+02;
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 3 AANNAGICVPCQINRGV 21
Db 333 ATGNVSGVCCDCHNTMG 351

RESULT 12
US-08-038-682-4
Sequence 4, Application US/08038682
Patent No. 5549897
GENERAL INFORMATION:
APPLICANT: BARENKAMP, STEPHEN J
APPLICANT: ST. GEME III, JOSEPH W
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: Shoemaker and Matlare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/038.682
FILING DATE: 16-MAR-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERKSTRESSER, JERRY W
REGISTRATION NUMBER: 22.651
REFERENCE/DOCKET NUMBER: 1038-293
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-038-682-4

Query Match 26.1%; Score 46; DB 1; Length 1477;

Best Local Similarity 40.0%; Pred. No. 3e+02;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

OY 5 NNAAGICVPCQINRGVSTNAGDL 29
Db 925 NNGTAENITGCVKLGNTNDGDL 949

RESULT 13
US-08-302-832-4
Sequence 4, Application US/08302832
Patent No. 5603938
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: Shoemaker and Matlare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302.832
FILING DATE: 16-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US pct/us93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22.651
REFERENCE/DOCKET NUMBER: 1038-404
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-302-832-4

Query Match 26.1%; Score 46; DB 1; Length 1477;
Best Local Similarity 40.0%; Pred. No. 3e+02;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

OY 5 NNAAGICVPCQINRGVSTNAGDL 29
Db 925 NNGTAENITGCVKLGNTNDGDL 949

RESULT 14
US-08-530-198-4
Sequence 4, Application US/08530198
Patent No. 5869065
GENERAL INFORMATION:
APPLICANT: BARENKAMP, STEPHEN J
APPLICANT: ST. GEME III, JOSEPH W
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS

TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Shoemaker and Mattare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/530,198
FILING DATE: 13-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERKSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: JWB-1186
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-530-198-4

Query Match 26.1%; Score 46; DB 2; Length 1477;
Best Local Similarity 40.0%; Pred. No. 3e+02;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 5 NNAAGICVPCQINRGVSTNAGDL 29
DB 925 NNGTAEINITGGVVKLGNTNDGL 949

RESULT 15
US-08-469-880-4
Sequence 4, Application US/08469880
Patent No. 5876733
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5876733-typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,880
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1

FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MIS:V9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-469-880-4

Query Match 26.1%; Score 46; DB 2; Length 1477;
Best Local Similarity 40.0%; Pred. No. 3e+02;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 5 NNAAGICVPCQINRGVSTNAGDL 29
DB 925 NNGTAEINITGGVVKLGNTNDGL 949

Search completed: March 6, 2001, 12:53:49
JOB time: 349 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:50:48 ; Search time 70.34 Seconds
(without alignments)
32.821 Million cell updates/sec

Title: US-09-196-161d-4
Perfect score: 176
Sequence: 1 PFANNAARGICVPCQINRVGVTNAGDLATLAT 34

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_66:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	176	100.0	395	2 A46031	Immobiliation sur
2	53	30.1	3712	2 S18253	laminin alpha-1 ch
3	52	29.5	101	2 A43262	hypothetical prote
4	52	29.5	303	2 B11638	cell division prot
5	50	28.4	261	2 S58496	IAI division - Ara
6	49.5	28.1	706	2 D81236	nitrogen regulatio
7	49	27.8	318	2 T41105	60s ribosomal prot
8	49	27.6	1394	2 T34061	hypothetical prote
9	48.5	27.6	706	2 E82009	probable two-compo
10	48	27.3	370	2 A72342	hypothetical prote
11	48	27.3	400	1 NOHSM	phosphopyruvate hy
12	48	27.3	846	2 A30889	integrin beta chain
13	47.5	27.0	393	2 S27881	beta-alanine synth
14	47	26.7	240	1 A70031	conserved hypotet
15	47	26.7	1790	1 MMF81	laminin beta-1 cha
16	46.5	26.4	1372	2 T25933	hypothetical prote
17	46.5	26.4	3512	2 T17121	CPY protein - midg
18	46	26.1	92	2 T30414	conotoxin homology
19	46	26.1	245	2 T03413	probable 1-aminoc
20	46	26.1	247	2 E75254	conserved hypotet
21	46	26.1	286	2 E83048	hypothetical prote
22	46	26.1	292	2 E82385	hypothetical prote
23	46	26.1	313	2 E82228	histidine ammonia-
24	46	26.1	511	2 T34371	hypothetical prote
25	46	26.1	609	2 T08050	probable ethylene
26	46	26.1	637	2 T08050	env polyprotein pr
27	46	26.1	843	1 H44001	high-molecular-wt
28	46	26.1	1477	2 B43855	laminin beta-1 cha
29	46	26.1	1786	1 MMHDB1	

30	46	26.1	2823	2 T23064	hypothetical prote
31	46	26.1	3102	2 T43291	laminin alpha chain
32	45	25.6	81	2 I48749	s-laminin - mouse
33	45	25.6	89	2 S61244	probable myristyla
34	45	25.6	180	2 T43451	hypothetical prote
35	45	25.6	184	2 C83373	hypothetical prote
36	45	25.6	252	2 PC1140	cellulase (Ec 3.2.
37	45	25.6	302	2 T15936	hypothetical prote
38	45	25.6	424	2 E75362	glutamate dehydrog
39	45	25.6	434	1 NOHUG	phosphopyruvate hy
40	45	25.6	434	2 A24742	phosphopyruvate hy
41	45	25.6	434	2 S10247	phosphopyruvate hy
42	45	25.6	434	2 JC1039	phosphopyruvate hy
43	45	25.6	434	2 T25040	hypothetical prote
44	45	25.6	463	2 A40013	phosphomannomutase
45	45	25.6	463	2 H82979	phosphomannomutase

ALIGNMENTS

```

RESULT 1
A46031
Immobiliation surface T-antigen precursor - Ichthyophthirius multifiliis (fragment)
C:Species: Ichthyophthirius multifiliis
C:Date: 29-Sep-1999 #sequence_revision 29-Sep-1999 #text_change 07-Dec-1999
C:Accession: A46031
R:Clark, T.G., McGraw, R.A., Dickerson, H.W.
Proc. Natl. Acad. Sci. U.S.A. 89, 6363-6367, 1992
A:Title: Developmental expression of surface antigen genes in the parasitic ciliate I
A:Reference number: A46031; MUID:92335298
A:Accession: A46031
A:Molecule type: mRNA; protein
A:Residues: 1-395 <CIA>
A:CROSS-references: GB:M92907; NID:q3628568; PIDN:ANC36158.1; PID:q3628569
A:Note: The authors translated the codon UUG for residue 350 as Ile
A:Note: sequence extracted from NCBI database (NCBIN:108734, NCBIPI:108735); the sequen
C:Genetics:
A:Genetic code: SGC5
C:Keywords: glycoprotein; surface antigen
F:2-395/Product: Immobiliation surface T-antigen #status experimental <Mat>
F:156,191,245,281/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 176; DB 2; Length 395;
Best Local Similarity 100.0%; Pred. No. 3.9e-17;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PFANNAARGICVPCQINRVGVTNAGDLATLAT 34
DB 48 PFANNAARGICVPCQINRVGVTNAGDLATLAT 81

RESULT 2
S18253
laminin alpha-1 chain precursor - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 16-Sep-1997 #sequence_revision 24-Jul-1997 #text_change 10-Dec-1999
C:Accession: S28399; S18253
R:Kusche-Gullberg, M.; Garlson, K.; Mackrell, A.J.; Fessler, L.I.; Fessler, J.H.
EMBO J. 11, 4519-4527, 1992
A:Title: Laminin A chain: expression during Drosophila development and genomic sequen
A:Reference number: S28399; MUID:93049203
A:Accession: S28399
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-3712 <KUS>
A:CROSS-references: GB:M96388; NID:q157799; PIDN:AAA28662.1; PID:q157800
R:Garlson, K.; Mackrell, A.J.; Fessler, J.H.
J. Biol. Chem. 266, 22898-22904, 1991
A:Title: Drosophila laminin A chain sequence, interspecies comparison, and domain str
A:Reference number: S18253; MUID:92078147
A:Accession: S18253

```

A:Molecule type: mRNA
A:Residues: 1762-3712 <GAR>
A:Cross-references: EMBL:M75882; NID:q157797; PIDD:AAA28661.1; PID:q157798
C:Genetics:
A:Gene: FlyBase:LaNa
A:Cross-references: FlyBase:FBgn0002526
C:Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like homology
C:Keywords: basement membrane; cell binding; coiled coil; disulfide bond; extracellular
F:273-330/Domain: laminin-type EGF-like homology <LEG>
F:333-400/Domain: laminin-type EGF-like homology <LE02>
F:541-584/Domain: laminin-type EGF-like homology <LEG1>
F:1776-2115/Domain: III <DOM3>
F:1776-1806/Domain: laminin-type EGF-like homology #status atypical <LEI>
F:1809-1856/Domain: laminin-type EGF-like homology <LE2>
F:1859-1914/Domain: laminin-type EGF-like homology <LE3>
F:1917-1967/Domain: laminin-type EGF-like homology <LE4>
F:1970-2014/Domain: laminin-type EGF-like homology <LE5>
F:2017-2061/Domain: laminin-type EGF-like homology <LE6>
F:2064-2109/Domain: laminin-type EGF-like homology <LE7>
F:2116-2697/Domain: I/II, heptad repeats <DOM2>
F:2698-3712/Domain: G <DOMG>
F:2698-2863/Domain: repeat G1 <RG1>
F:2864-3048/Domain: repeat G2 <RG2>
F:3049-3223/Domain: repeat G3 <RG3>
F:3079-3200/Domain: laminin G repeat homology <LG3>
F:3334-3528/Domain: repeat G4 <RG4>
F:3529-3712/Domain: repeat G5 <RG5>
F:1847,1850,1943,2024,2196,2215,2267,2301,2323,2482,2524,2538,2569,2699,2720,2890,2938,3011,3049,3112,3147,3176,3200,3223,3247,3270,3294,3318,3341,3364,3387,3410,3433,3456,3479,3502,3525,3548,3571,3594,3617,3640,3663,3686,3709,3732,3755,3778,3801,3824,3847,3870,3893,3916,3939,3962,3985,4008,4031,4054,4077,4100,4123,4146,4169,4192,4215,4238,4261,4284,4307,4330,4353,4376,4399,4422,4445,4468,4491,4514,4537,4560,4583,4606,4629,4652,4675,4698,4721,4744,4767,4790,4813,4836,4859,4882,4905,4928,4951,4974,4997,5020,5043,5066,5089,5112,5135,5158,5181,5204,5227,5250,5273,5296,5319,5342,5365,5388,5411,5434,5457,5480,5503,5526,5549,5572,5595,5618,5641,5664,5687,5710,5733,5756,5779,5802,5825,5848,5871,5894,5917,5940,5963,5986,6009,6032,6055,6078,6101,6124,6147,6170,6193,6216,6239,6262,6285,6308,6331,6354,6377,6400,6423,6446,6469,6492,6515,6538,6561,6584,6607,6630,6653,6676,6699,6722,6745,6768,6791,6814,6837,6860,6883,6906,6929,6952,6975,6998,7021,7044,7067,7090,7113,7136,7159,7182,7205,7228,7251,7274,7297,7320,7343,7366,7389,7412,7435,7458,7481,7504,7527,7550,7573,7596,7619,7642,7665,7688,7711,7734,7757,7780,7803,7826,7849,7872,7895,7918,7941,7964,7987,8010,8033,8056,8079,8102,8125,8148,8171,8194,8217,8240,8263,8286,8309,8332,8355,8378,8401,8424,8447,8470,8493,8516,8539,8562,8585,8608,8631,8654,8677,8700,8723,8746,8769,8792,8815,8838,8861,8884,8907,8930,8953,8976,8999,9022,9045,9068,9091,9114,9137,9160,9183,9206,9229,9252,9275,9298,9321,9344,9367,9390,9413,9436,9459,9482,9505,9528,9551,9574,9597,9620,9643,9666,9689,9712,9735,9758,9781,9804,9827,9850,9873,9896,9919,9942,9965,9988,10011,10034,10057,10080,10103,10126,10149,10172,10195,10218,10241,10264,10287,10310,10333,10356,10379,10402,10425,10448,10471,10494,10517,10540,10563,10586,10609,10632,10655,10678,10701,10724,10747,10770,10793,10816,10839,10862,10885,10908,10931,10954,10977,11000,11023,11046,11069,11092,11115,11138,11161,11184,11207,11230,11253,11276,11299,11322,11345,11368,11391,11414,11437,11460,11483,11506,11529,11552,11575,11598,11621,11644,11667,11690,11713,11736,11759,11782,11805,11828,11851,11874,11897,11920,11943,11966,11989,12012,12035,12058,12081,12104,12127,12150,12173,12196,12219,12242,12265,12288,12311,12334,12357,12380,12403,12426,12449,12472,12495,12518,12541,12564,12587,12610,12633,12656,12679,12702,12725,12748,12771,12794,12817,12840,12863,12886,12909,12932,12955,12978,13001,13024,13047,13070,13093,13116,13139,13162,13185,13208,13231,13254,13277,13300,13323,13346,13369,13392,13415,13438,13461,13484,13507,13530,13553,13576,13599,13622,13645,13668,13691,13714,13737,13760,13783,13806,13829,13852,13875,13898,13921,13944,13967,13990,14013,14036,14059,14082,14105,14128,14151,14174,14197,14220,14243,14266,14289,14312,14335,14358,14381,14404,14427,14450,14473,14496,14519,14542,14565,14588,14611,14634,14657,14680,14703,14726,14749,14772,14795,14818,14841,14864,14887,14910,14933,14956,14979,15002,15025,15048,15071,15094,15117,15140,15163,15186,15209,15232,15255,15278,15301,15324,15347,15370,15393,15416,15439,15462,15485,15508,15531,15554,15577,15600,15623,15646,15669,15692,15715,15738,15761,15784,15807,15830,15853,15876,15899,15922,15945,15968,15991,16014,16037,16060,16083,16106,16129,16152,16175,16198,16221,16244,16267,16290,16313,16336,16359,16382,16405,16428,16451,16474,16497,16520,16543,16566,16589,16612,16635,16658,16681,16704,16727,16750,16773,16796,16819,168

```

Nature 396, 15-140, 1998
A>Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499
A:Accession: B11638
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-303 <AND>
A:Cross-references: GB:A235273; GB:A235269; NID:g3861237; PID:CAA15202.1; PID:g3866
A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: ftsY; RP775
C:Superfamily: docking protein

Query Match      29.5%; Score 52; DB 2; Length 303;
Best Local Similarity 34.4%; Pred. No. 7.9;
Matches 11; Conservative 4; Mismatches 17; Indels 0; Gaps 0;

QY      1 PFAANNARGICVPCQINRVGSVTNAGDLATL 32
       .|||::|||::|||::|||::|||::|||::|||
Db      93 PFTESENKLITLLVCGVNGVGKTTTIGKLSAL 124

RESULT      5
S58496
TAAL protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 29-Nov-1995 #sequence_revision 01-Mar-1996 #text_change 21-Jul-2000
C:Accession: S58496; T000963
R:Abel, S.; Nguyen, M.D.; Theologis, A., J. Mol. Biol. 251, 533-549, 1995
A>Title: The ps-TAA4/5-like family of early auxin-inducible mRNAs in Arabidopsis thal
A:Reference number: S58491; MUID:95387393
A:Accession: S58496
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-261 <ABR>
A:Cross-references: EMBL:U18412; NID:g972922; PID:NAC49051.1; PID:g972923
R:Vysockakata, V.S.; Osborne, B.I.; Schwartz, J.R.; Toriumi, M.; Kwan, A.; Yu, G.; Ojima
K.; Feng, J.; Kim, C.; Kurtz, D.; Li, Y.; Palm, C.J.; Shinn, P.; Sun, H.; Davis, R.W.
submitted to the EMBL Data Library, May 1998
A:Description: Arabidopsis thaliana chromosome 1 BAC F20022 complete sequence.
A:Reference number: Z14214
A:Accession: T00963
A>Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-261 <VYS>
A:Cross-references: EMBL:AC002411; NID:g2570223; PID:g3142299
C:Genetics:
A:Map position: I
A:Introns: 91/2; 195/1; 239/2
A>Note: F20D22.13
C:Superfamily: auxin-induced protein aux28

Query Match      28.4%; Score 50; DB 2; Length 261;
Best Local Similarity 42.9%; Pred. No. 13;
Matches 15; Conservative 5; Mismatches 11; Indels 4; Gaps 2;

QY      1 PFAANNARGICV--PCQINRVGSVTN-AGDDLAR 31
       |||::|||::|||::|||::|||::|||::|||
Db      83 PAAMSNATROYAVGWPUPRTYRKINSLVNAOKSLAR 117

RESULT      6
DB1236
nitrogen regulation protein NtrY, probable NMB0114 [imported] - Neisseria meningitidis
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
C:Accession: DB1236
R:Reitelstein, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.;
Lin, H.; Qin, H.; Vanthethanan, J.; Gill, J.; Scarlato, V.; Mastignani, V.; Plizsa, M.;

```

Science 287, 1809-1815, 2000
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappunli, R.; V
 A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain Mc58.
 A:Reference number: A61000; MUID:2017575
 A:Accession: D81236
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-706 <TEF>
 A:Cross-references: GB:AE002370; GB:AE002098; NID:g7225327; PID:NAAF40573.1; PID:g722533
 A:Experimental source: serogroup B, strain MMD58
 C:Genetics:
 A:Gene: MMR0114

A:Accession: A28783
A:Molecule type: mRNA
A:Residues: 1-1790 <MON1>
A:Cross-references: EMBL:M19525
R:Montell, D.J.; Goodman, C.S.
submitted to the EMBL Data Library, June 1988
A:Description: Drosophila substrate adhesion molecule: sequence of laminin BI chain reve
A:Reference number: S14462
A:Accession: S14462
A:Molecule type: mRNA
A:Residues: 1-667, 'L', 669-725, 'VT', 728-947, 950-1790 <MON2>
A:Cross-references: EMBL:M19525; NID:g157801; PIDN:AAA28663.1; PID:g157802
C:GeneLocs:
A:Gene: Lamb1
A:Cross-references: FlyBase:FBgn0002527
A:Map position: 2L:28D
C:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin
C:Function:
A:Description: Interact with cells and with other basement membrane proteins to promote
C:Superfamily: laminin beta-1 chain; laminin-type EGF-like homology
C:Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellular
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-1790/Product: laminin beta-1 chain #status predicted <MAT>
F:27-288/Domain: VI <DOM6>
F:289-561/Domain: V <DOM5>
F:290-354/Domain: laminin-type EGF-like homology <LE01>
F:357-417/Domain: laminin-type EGF-like homology <LE02>
F:420-477/Domain: laminin-type EGF-like homology <LE03>
F:480-528/Domain: laminin-type EGF-like homology <LE04>
F:531-561/Domain: laminin-type EGF-like homology #status atypical <LE05>
F:562-789/Domain: IV <DOM4>
F:643-645/Region: cell attachment (R-G-D) motif
F:790-1189/Domain: III <DOM3>
F:791-836/Domain: laminin-type EGF-like homology <LE06>
F:839-882/Domain: laminin-type EGF-like homology <LE07>
F:886-932/Domain: laminin-type EGF-like homology <LE08>
F:935-990/Domain: laminin-type EGF-like homology <LE09>
F:968-972/Region: cell attachment #status predicted
F:993-1042/Domain: laminin-type EGF-like homology <LE10>
F:1045-1093/Domain: laminin-type EGF-like homology <LE11>
F:1096-1141/Domain: laminin-type EGF-like homology <LE12>
F:1144-1188/Domain: laminin-type EGF-like homology <LE13>
F:1190-1407/Domain: II <DOM2>
F:1408-1434/Domain: alpha <ALP>
F:1435-1790/Domain: I <DOM1>
F:51-56/Disulfide bonds: #status predicted
F:140,203,234,489,593,1053,1248,1303,1332,1343,1475,1495,1517,1583,1646,1705/Binding sit
F:1191,1194,1788/Disulfide bonds: Interchain #status predicted

Query Match 26.7%; Score 47; DB 1; Length 1790;
Best local similarity 43.8%; Pred. No. 2.2e+02;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 12 CVPCQINRVGSVTNAG 27
| | | | : | : | : |
DB 477 CEPCTCNPLGTLNNSG 492

Search completed: March 6, 2001, 12:50:50
Job time: 171 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:54:33 ; Search time 38.83 Seconds
(without alignments)
28.277 Million cell updates/sec

Title: US-09-196-161D-4
Perfect score: 176
Sequence: 1 PFANNANRAGICVPCQINRVSVTNAGDIATLAT 34

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 88757 seqs, 32294092 residues
Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	30.1	3712	1	LMA_DROME
2	52	29.5	303	1	FTSY_RICPR
3	48	27.3	400	1	ENO_HALMA
4	48	27.3	775	1	LYS4_EMENI
5	48	27.3	846	1	ITBX_DROME
6	47.5	27.0	393	1	BUP_RAT
7	47.5	27.0	3110	1	LMA2_HUMAN
8	47	26.7	1790	1	LMB1_DROME
9	46.5	26.4	452	1	PR11_PLAFK
10	46	26.1	433	1	ENO_HOMGA
11	46	26.1	843	1	ENO_HV1Y2
12	46	26.1	1786	1	LMB1_HUMAN
13	45	25.6	433	1	ENOG_HUMAN
14	45	25.6	433	1	ENOG_MOUSE
15	45	25.6	433	1	ENOG_RAT
16	45	25.6	433	1	ENO_CAEEL
17	45	25.6	434	1	ENOG_CHICK
18	45	25.6	462	1	ALGC_PSEAE
19	45	25.6	518	1	VIL_HPV21
20	45	25.6	546	1	VIL_HPV19
21	45	25.6	775	1	VP4_NOTHP
22	45	25.6	1157	1	C9CA_BACTO
23	45	25.6	1475	1	N153_HUMAN
24	45	25.6	1786	1	LMB1_MOUSE
25	45	25.6	1798	1	LMB2_HUMAN
26	45	25.6	1799	1	LMB2_MOUSE
27	45	25.6	1801	1	LMB2_RAT
28	45	25.6	2569	1	LMA3_MOUSE
29	44.5	25.3	302	1	COBK_METH
30	44.5	25.3	344	1	DHBI_MOUSE
31	44.5	25.3	344	1	DHBI_RAT
32	44.5	25.3	454	1	PR11_SCHPO
33	44	25.0	322	1	TF3A_ICMPU

34	44	25.0	352	1	MODC_ECOLI	P09833 escherichia
35	44	25.0	460	1	TUL3_MOUSE	O08413 mus musculu
36	44	25.0	598	1	P2CD_MOUSE	O94267 mus musculu
37	44	25.0	605	1	P2CD_HUMAN	O15297 homo sapien
38	44	25.0	646	1	FATP_RAT	P97849 rattus norv
39	44	25.0	704	1	Y590_METUA	O56010 methanococc
40	44	25.0	811	1	FS22_DROME	P34083 drosophilla
41	44	25.0	873	1	FS21_DROME	P34082 drosophilla
42	44	25.0	1285	1	SL17_ENTHI	P23502 entameeba h
43	44	25.0	3106	1	LMA2_MOUSE	O60675 mus musculu
44	43.5	24.7	193	1	YNES_BACSU	O45064 bacillus su
45	43.5	24.7	261	1	YV00_MYCLE	O49741 mycobacteri

ALIGNMENTS

RESULT 1
LMA_DROME STANDARD: PRT: 3712 AA.
AC 000174:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE LAMININ ALPHA CHAIN PRECURSOR.
GN LANA OR LAMA.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93049203; PubMed=1425586;
RA Kusche-Gullberg M., Garrison K., Mackrell A.J., Fessler L.I.,
RA Fessler J.H.;
RT "Laminin A chain: expression during Drosophila development and
RT genomic sequence.";
RL EMBO J. 11:4519-4527(1992).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE-EMBRYO;
RX MEDLINE=94036678; PubMed=8223265;
RA Henchcliffe C., Garcia-Alonso L., Tang J., Goodman C.S.;
RT "Genetic analysis of laminin A reveals diverse functions during
RT morphogenesis in Drosophila.";
RL Development 118:325-337(1993).
[3]
RP SEQUENCE OF 1762-3712 FROM N.A.
RX MEDLINE=92078147; PubMed=1744083;
RA Garrison K., Mackrell A.J., Fessler J.H.;
RT "Drosophila laminin A chain sequence, interspecies comparison, and
RT domain structure of a major carboxyl portion.";
RL J. Biol. Chem. 266:22899-22904(1991).
- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
- FUNCTION: DIVERSE FUNCTIONS DURING MORPHOGENESIS IN DROSOPHILA.
COMPLETE LOSS-OF-FUNCTION MUTATIONS LEAD TO LATE EMBRYONIC
LETHALITY. CERTAIN PARTIAL LOSS-OF-FUNCTION MUTATIONS GIVE RAISE
TO ESCAPER ADULTS, WHICH HAVE ROUGH EYES ASSOCIATED WITH CHANGES
IN CELL FATE AND PATTERN, MISSHAPEN LEGS AND DEFECTS IN WING
STRUCTURE.
- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
MEMBRANES (MAJOR COMPONENT).
- TISSUE SPECIFICITY: NEWLY FORMED MESODERM AND LATER PROMINENTLY
EXPRESSED IN HEMOCYTES, WHICH ALSO SYNTHESIZE COLLAGEN IV.
- DEVELOPMENTAL STAGE: DURING MORPHOGENESIS, MOSTLY IN EMBRYO
DEVELOPMENT AT 10-12 HOURS.

CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -1- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -1- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV (DOMAIN IV'' IS NOT
CC SIMILAR TO LAMININ DOMAIN IV).
CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M96388; AAA28662.1; -;
DR EMBL: L07288; AAC37178.1; -;
DR EMBL: M73882; AAA28661.1; -;
DR HSSP: P02468; 1TLE.
DR FLYBASE: FBgn0002526; Lana.
DR INTERPRO: IPR000034; -;
DR INTERPRO: IPR0000561; -;
DR INTERPRO: IPR001791; -;
DR INTERPRO: IPR001886; -;
DR INTERPRO: IPR002049; -;
DR PFAM: PF00052; laminin_B; 1.
DR PFAM: PF00053; laminin_EGF; 20.
DR PFAM: PF00054; laminin_G; 5.
DR PFAM: PF00055; laminin_Nterm; 1.
DR PRINTS: PR00011; EGFLAMININ.
DR PROSITE: PS00022; EGF_1; 17.
DR PROSITE: PS01186; EGF_2; 5.
DR PROSITE: PS01248; LAMININ_Type_EGF; 19.
DR Glycoprotein: Basement membrane; Extracellular matrix; Coiled coil;
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
FT SIGNAL 1 22
FT CHAIN 23 3712 LAMININ ALPHA CHAIN.
FT DOMAIN 25 272 LAMININ N-TERMINAL (DOMAIN VI).
FT DOMAIN 273 815 10.5 X LAMININ EGF-LIKE REPEATS (DOMAIN
FT V).
FT DOMAIN 273 332 LAMININ EGF-LIKE 1.
FT DOMAIN 333 402 LAMININ EGF-LIKE 2.
FT DOMAIN 403 447 LAMININ EGF-LIKE 3.
FT DOMAIN 448 494 LAMININ EGF-LIKE 4.
FT DOMAIN 495 540 LAMININ EGF-LIKE 5.
FT DOMAIN 541 586 LAMININ EGF-LIKE 6.
FT DOMAIN 587 631 LAMININ EGF-LIKE 7.
FT DOMAIN 632 676 LAMININ EGF-LIKE 8.
FT DOMAIN 677 731 LAMININ EGF-LIKE 9.
FT DOMAIN 732 784 LAMININ EGF-LIKE 10.
FT DOMAIN 785 815 LAMININ EGF-LIKE 11 (INCOMPLETE).
FT DOMAIN 816 1374 DOMAIN IV''.
FT DOMAIN 1375 1574 4.5 X LAMININ EGF-LIKE REPEATS (DOMAIN
FT III B).
FT DOMAIN 1375 1420 LAMININ EGF-LIKE 12.
FT DOMAIN 1421 1465 LAMININ EGF-LIKE 13.
FT DOMAIN 1466 1513 LAMININ EGF-LIKE 14.
FT DOMAIN 1514 1564 LAMININ EGF-LIKE 15.
FT DOMAIN 1565 1574 LAMININ EGF-LIKE 16 (N-TERMINAL).
FT DOMAIN 1575 1775 LAMININ DOMAIN IV (DOMAIN IV).
FT DOMAIN 1776 2111 6.5 X LAMININ EGF-LIKE REPEATS (DOMAIN
FT III A).
FT DOMAIN 1776 1808 LAMININ EGF-LIKE 16 (C-TERMINAL).
FT DOMAIN 1809 1858 LAMININ EGF-LIKE 17.
FT DOMAIN 1859 1916 LAMININ EGF-LIKE 18.
FT DOMAIN 1917 1969 LAMININ EGF-LIKE 19.
FT DOMAIN 1970 2016 LAMININ EGF-LIKE 20.
FT DOMAIN 2017 2063 LAMININ EGF-LIKE 21.
FT DOMAIN 2064 2111 LAMININ EGF-LIKE 22.
FT DOMAIN 2112 2697 DOMAIN II AND I.
FT DOMAIN 2698 3712 5 X LAMININ G-LIKE REPEATS (DOMAIN G).

FT DOMAIN 2698 2862 LAMININ G-LIKE 1.
FT DOMAIN 2863 3048 LAMININ G-LIKE 2.
FT DOMAIN 3049 3323 LAMININ G-LIKE 3.
FT DOMAIN 3323 3296 POLY-THR.
FT DOMAIN 3334 3528 LAMININ G-LIKE 4.
FT DOMAIN 3529 3712 LAMININ G-LIKE 5.
FT DOMAIN 3712 2249 COILED COIL (POTENTIAL).
FT DOMAIN 2249 2301 COILED COIL (POTENTIAL).
FT DOMAIN 2301 2321 COILED COIL (POTENTIAL).
FT DOMAIN 2321 2450 COILED COIL (POTENTIAL).
FT DOMAIN 2450 2541 COILED COIL (POTENTIAL).
FT DOMAIN 2541 2676 BY SIMILARITY.
FT DOMAIN 2676 273 BY SIMILARITY.
FT DOMAIN 273 282 BY SIMILARITY.
FT DOMAIN 282 298 BY SIMILARITY.
FT DOMAIN 298 307 BY SIMILARITY.
FT DOMAIN 307 330 BY SIMILARITY.
FT DOMAIN 330 342 BY SIMILARITY.
FT DOMAIN 342 367 BY SIMILARITY.
FT DOMAIN 367 379 BY SIMILARITY.
FT DOMAIN 379 400 BY SIMILARITY.
FT DOMAIN 400 414 BY SIMILARITY.
FT DOMAIN 414 421 BY SIMILARITY.
FT DOMAIN 421 432 BY SIMILARITY.
FT DOMAIN 432 445 BY SIMILARITY.
FT DOMAIN 445 460 BY SIMILARITY.
FT DOMAIN 460 468 BY SIMILARITY.
FT DOMAIN 468 479 BY SIMILARITY.
FT DOMAIN 479 492 BY SIMILARITY.
FT DOMAIN 492 507 BY SIMILARITY.
FT DOMAIN 507 514 BY SIMILARITY.
FT DOMAIN 514 525 BY SIMILARITY.
FT DOMAIN 525 538 BY SIMILARITY.
FT DOMAIN 538 541 BY SIMILARITY.
FT DOMAIN 541 553 BY SIMILARITY.
FT DOMAIN 553 560 BY SIMILARITY.
FT DOMAIN 560 571 BY SIMILARITY.
FT DOMAIN 571 584 BY SIMILARITY.
FT DOMAIN 584 599 BY SIMILARITY.
FT DOMAIN 599 605 BY SIMILARITY.
FT DOMAIN 605 616 BY SIMILARITY.
FT DOMAIN 616 629 BY SIMILARITY.
FT DOMAIN 629 644 BY SIMILARITY.
FT DOMAIN 644 650 BY SIMILARITY.
FT DOMAIN 650 661 BY SIMILARITY.
FT DOMAIN 661 674 BY SIMILARITY.
FT DOMAIN 674 691 BY SIMILARITY.
FT DOMAIN 691 700 BY SIMILARITY.
FT DOMAIN 700 711 BY SIMILARITY.
FT DOMAIN 711 729 BY SIMILARITY.
FT DOMAIN 729 746 BY SIMILARITY.
FT DOMAIN 746 753 BY SIMILARITY.
FT DOMAIN 753 764 BY SIMILARITY.
FT DOMAIN 764 782 BY SIMILARITY.
FT DOMAIN 782 1387 BY SIMILARITY.
FT DOMAIN 1387 1394 BY SIMILARITY.
FT DOMAIN 1394 1405 BY SIMILARITY.
FT DOMAIN 1405 1429 BY SIMILARITY.
FT DOMAIN 1429 1436 BY SIMILARITY.
FT DOMAIN 1436 1447 BY SIMILARITY.
FT DOMAIN 1447 1463 BY SIMILARITY.
FT DOMAIN 1463 1480 BY SIMILARITY.
FT DOMAIN 1480 1487 BY SIMILARITY.
FT DOMAIN 1487 1498 BY SIMILARITY.
FT DOMAIN 1498 1511 BY SIMILARITY.
FT DOMAIN 1511 1526 BY SIMILARITY.
FT DOMAIN 1526 1533 BY SIMILARITY.
FT DOMAIN 1533 1544 BY SIMILARITY.
FT DOMAIN 1544 1562 BY SIMILARITY.
FT DOMAIN 1562 1874 BY SIMILARITY.
FT DOMAIN 1874 1885 BY SIMILARITY.
FT DOMAIN 1885 1896 BY SIMILARITY.
FT DOMAIN 1896 1914 BY SIMILARITY.
FT DOMAIN 1914 1931 BY SIMILARITY.
FT DOMAIN 1931 1938 BY SIMILARITY.
FT DOMAIN 1938 1950 BY SIMILARITY.


```

RESULT      6
BUP_RAT     STANDARD:      PRT:      393 AA.
ID          003248:
AC          01-OCT-1993 (Rel. 27, Created)
DT          01-OCT-1993 (Rel. 27, Last sequence update)
DT          15-JUL-1999 (Rel. 38, Last annotation update)
DE          BETA-UREIDOPROPIONASE (EC 3.5.1.6) (BETA-ALANINE SYNTHASE)
DE          (N-CARBAMOYL-BETA-ALANINE AMIDOHYDROLASE).
OS          Rattus norvegicus (Rat).
OC          Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN          (1)
RN          SEQUENCE FROM N.A., AND SEQUENCE OF 202-212.
RP          TISSUE=LIVER;
RC          MEDLINE=93194866; PubMed=8449931;
RA          Kvalnes-Krick K.L., Traut T.W.;
RT          "Cloning, sequencing, and expression of a cDNA encoding beta-alanine
RT          synthase from rat liver.";
RL          J. Biol. Chem. 268:5686-5693(1993).
CC          -1- CATALYTIC ACTIVITY: N-CARBAMOYL-BETA-ALANINE + H(2)O =
CC          BETA-ALANINE + CO(2) + NH(3).
CC          -1- COFACTOR: BINDS TWO ZINC IONS.
CC          -1- ENZYME REGULATION: ALLOSTERIC ENZYME WITH POSITIVE COOPERATIVITY
CC          TOWARD THE SUBSTRATE N-CARBAMOYL-BETA-ALANINE.
CC          -1- PATHWAY: THIRD AND FINAL STEP IN THE CATABOLISM OF THE
CC          PYRIMIDINE BASES, URACIL OR THYMINE, TO PRODUCE BETA-ALANINE OR
CC          2-METHYL-BETA-ALANINE.
CC          -1- SUBUNIT: IN THE ABSENCE OF LIGANDS, THE ENZYME EXISTS AS A STABLE
CC          HOMODIMER, ALTHOUGH THIS IS NOT THE MOST ACTIVE FORM OF THE
CC          ENZYME. THIS NATIVE HEXAMER DISSOCIATES TO AN INACTIVE TRIMER IN
CC          RESPONSE TO THE PRODUCT, BETA-ALANINE, OR ASSOCIATES TO THE MORE
CC          ACTIVE HOMODODECAMER IN RESPONSE TO THE SUBSTRATE.
CC          -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC          -----
CC          This SWISS-PROT entry is copyright. It is produced through a collaboration
CC          between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC          the European Bioinformatics Institute. There are no restrictions on its
CC          use by non-profit institutions as long as its content is in no way
CC          modified and this statement is not removed. Usage by and for commercial
CC          entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC          or send an email to license@isb-sib.ch).
CC          -----
DR          EMBL; M97662; AAA40804.1; .
DR          PIR; S27881; S27881.
DR          PIR; A46624; A46624.
DR          INTERPRO; IPR003010; .
DR          PIR; PF00795; CN_hydrolase; 1.
KW          Hydrolase; zinc.
FT          MOD_RES      ?1      ?1      BLOCKED.
FT          METAL        97      97      ZINC 1 (POTENTIAL).
FT          METAL        101     101     ZINC 1 (POTENTIAL).
FT          METAL        158     158     ZINC 1 (POTENTIAL).
FT          METAL        280     280     ZINC 2 (POTENTIAL).
FT          METAL        293     293     ZINC 2 (POTENTIAL).
FT          METAL        297     297     ZINC 2 (POTENTIAL).
SQ          SEQUENCE      393 AA; 44042 MW; 723E5AF5B01E3AAD CRC64;
Query Match      27.0%; Score 47.5; DB 1; Length 393;
Best Local Similarity 50.0%; Pred. No. 21;
Matches 11; Conservative 1; Mismatches 9; Indels 1; Gaps 1;

```

```

DT          01-MAR-1992 (Rel. 21, Created)
DT          01-NOV-1997 (Rel. 35, Last sequence update)
DT          15-JUL-1999 (Rel. 38, Last annotation update)
DE          LAMININ ALPHA-2 CHAIN PRECURSOR (LAMININ M CHAIN) (MEROSIN HEAVY
DE          CHAIN)
OS          Homo sapiens (Human).
OC          Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC          Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN          (1)
RN          SEQUENCE FROM N.A.
RP          TISSUE=PLACENTA;
RC          MEDLINE=94124633; PubMed=8294519;
RA          Virolenaho R., Nissinen M., Salino K., Byers M., Eddy R.,
RA          Hyytonen H., Shovs T.B., Sariola H., Engvall E., Tryggvason K.;
RT          "Human laminin M chain (merosin): complete primary structure,
RT          chromosomal assignment, and expression of the M and A chain in human
RT          fetal tissues.";
RL          J. Cell Biol. 124:381-394(1994).
RN          (2)
RN          SEQUENCE OF 1981-3110 FROM N.A., AND PARTIAL SEQUENCE.
RP          TISSUE=PLACENTA;
RC          MEDLINE=90238994; PubMed=2185464;
RA          Ehrig K., Leivo I., Argaves W.S., Ruoslahti E., Engvall E.;
RT          "Merosin, a tissue-specific basement membrane protein, is a
RT          laminin-like protein.";
RL          Proc. Natl. Acad. Sci. U.S.A. 87:3264-3268(1990).
RN          (3)
RN          VARIANTS GLN-545, HIS-619; LEU-919; HIS-2586 AND LYS-2614.
RA          Panicker S.G., Mendell J.T., Chen L., Feng B., Sahenk Z.,
RA          Marzluft G.A., Amato A.A., Mendell J.R.;
RT          "Novel single base polymorphisms and rare sequence variants in
RT          the laminin 2-chain coding region detected by RNA/SSCP analysis.";
RL          Hum. Mutat. 13:174-174(1999).
RN          (4)
RN          ERRATUM.
RA          Panicker S.G., Mendell J.T., Chen L., Feng B., Sahenk Z.,
RA          Marzluft G.A., Amato A.A., Mendell J.R.;
RL          Hum. Mutat. 13:340-340(1999).
CC          -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR. LAMININ
CC          IS THOUGHT TO MEDIANE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
CC          CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
CC          WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
CC          -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
CC          DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
CC          TO EACH OTHER BY DISULFIDE BONDS. INTO A CROSS-SHAPED MOLECULE
CC          COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
CC          THE ALPHA-2 CHAIN IS A SUBUNIT OF LAMININ-2 (MEROSIN) AND LAMININ-
CC          4 (S-MEROSIN).
CC          -1- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
CC          MEMBRANES (MAJOR COMPONENT).
CC          -1- TISSUE SPECIFICITY: PLACENTA, STRIATED MUSCLE, PERIPHERAL NERVE,
CC          CARDIAC MUSCLE, PANCREAS, LUNG, SPLEEN, KIDNEY, ADRENAL GLAND,
CC          SKIN, TESTIS, MENINGES, CHOROID PLEXUS, AND SOME OTHER REGIONS OF
CC          THE BRAIN; NOT IN LIVER, THYMUS AND BONE.
CC          -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC          WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC          -1- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
CC          -1- DISEASE: DEFECTS IN LAMA2 ARE THE CAUSE OF MEROSIN-DEFICIENT
CC          CONGENITAL MUSCULAR DYSTROPHY (MCDMD).
CC          -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC          -1- SIMILARITY: CONTAINS 17 LAMININ EGF-LIKE DOMAINS.
CC          -1- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
CC          -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC          -----
CC          This SWISS-PROT entry is copyright. It is produced through a collaboration
CC          between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC          the European Bioinformatics Institute. There are no restrictions on its
CC          use by non-profit institutions as long as its content is in no way
CC          modified and this statement is not removed. Usage by and for commercial
CC          entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC          or send an email to license@isb-sib.ch).

```

[illegible]

FT	DISULFID	759	773	BY SIMILARITY.
FT	DISULFID	776	785	BY SIMILARITY.
FT	DISULFID	768	804	BY SIMILARITY.
FT	DISULFID	807	822	BY SIMILARITY.
FT	DISULFID	809	832	BY SIMILARITY.
FT	DISULFID	835	844	BY SIMILARITY.
FT	DISULFID	847	862	BY SIMILARITY.
FT	DISULFID	865	879	BY SIMILARITY.
FT	DISULFID	867	886	BY SIMILARITY.
FT	DISULFID	889	898	BY SIMILARITY.
FT	DISULFID	901	915	BY SIMILARITY.
FT	DISULFID	918	930	BY SIMILARITY.
FT	DISULFID	920	937	BY SIMILARITY.
FT	DISULFID	939	948	BY SIMILARITY.
FT	DISULFID	951	964	BY SIMILARITY.
FT	DISULFID	967	979	BY SIMILARITY.
FT	DISULFID	969	985	BY SIMILARITY.
FT	DISULFID	987	996	BY SIMILARITY.
FT	DISULFID	989	1011	BY SIMILARITY.
FT	DISULFID	1014	1023	BY SIMILARITY.
FT	DISULFID	1016	1030	BY SIMILARITY.
FT	DISULFID	1032	1041	BY SIMILARITY.
FT	DISULFID	1044	1057	BY SIMILARITY.
FT	DISULFID	1060	1072	BY SIMILARITY.
FT	DISULFID	1062	1079	BY SIMILARITY.
FT	DISULFID	1081	1090	BY SIMILARITY.
FT	DISULFID	1093	1103	BY SIMILARITY.
FT	DISULFID	1420	1429	BY SIMILARITY.
FT	DISULFID	1422	1436	BY SIMILARITY.
FT	DISULFID	1439	1448	BY SIMILARITY.
FT	DISULFID	1451	1466	BY SIMILARITY.
FT	DISULFID	1469	1484	BY SIMILARITY.
FT	DISULFID	1471	1494	BY SIMILARITY.
FT	DISULFID	1497	1506	BY SIMILARITY.
FT	DISULFID	1509	1524	BY SIMILARITY.
FT	DISULFID	1527	1539	BY SIMILARITY.
FT	DISULFID	1529	1546	BY SIMILARITY.
FT	DISULFID	1548	1557	BY SIMILARITY.
FT	DISULFID	1560	1571	BY SIMILARITY.
FT	DISULFID	1574	1574	INTERCHAIN (PROBABLE).
FT	DISULFID	1578	1578	INTERCHAIN (PROBABLE).
FT	CARBOHYD	55	55	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	89	89	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	303	303	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	363	363	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	380	380	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	470	470	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	746	746	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1061	1061	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1597	1597	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1614	1614	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1700	1700	N-LINKED (GLCNAC. . .) (POTENTIAL).
Queue Match				
Best Local Similarity		27.0%;	Score 47.5;	DB 1;
Matches		10;	Conservative	1;
Mismatches		4;	Indels	1;
Gaps		1;		
QY	7	AARGICVPCQIMRGS	22	
DB	960	SARG-CVPCNCNSFGS	974	
RESULT 8				
LMBI_DROME	STANDARD.	PRT:	1790	AA.
AC	P11046; Q26328; Q9X2T4; Q9YLW6;			
DT	01-JUL-1989 (Rel. 11, Created)			
DT	01-OCT-2000 (Rel. 40, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	LAMININ BETA-1 CHAIN PRECURSOR (LAMININ BI CHAIN).			
GN	LAMBI OR LAMBI.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-CANTON-S;
 RX MEDLINE=94000382; PubMed=8397815;
 RA Gow C.-H., Chang H.-Y., Lin C.-J., Chang T.-W., Hui C.-F.;
 RT "Analysis of the Drosophila gene for the laminin BI chain."
 RL DNA Cell Biol. 12:573-587(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=88210471; PubMed=3365769;
 RA Montell D.J., Goodman C.S.;
 RT "Drosophila substrate adhesion molecule: sequence of laminin BI chain
 RT reveals domains of homology with mouse."
 RL Cell 53:463-473(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RC MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkov A., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadien L.B., Davies P.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos K., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heilmann T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
 RA Jajali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodside T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zavatt J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR. LAMININ
 CC IS THOUGHT TO MEDATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
 CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
 CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
 CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
 CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
 CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
 CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
 CC COMPONENT).
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -1- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
 CC -1- SIMILARITY: CONTAINS I LAMININ N-TERMINAL DOMAIN (DOMAIN VI).

CC -1- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC -----
 DR EMBL: M95811; AA019752.1; -;
 DR EMBL: M19525; AA28663.1; -;
 DR EMBL: AE003618; AAF52563.1; -;
 DR PIR: A28783; MMFEP1.
 DR HSSP: P02468; TITLE.
 DR FLYBASE: FBgn0002527; Lmb1.
 DR INTERPRO: IPR000561; -;
 DR INTERPRO: IPR001886; -;
 DR INTERPRO: IPR002049; -;
 DR PFAM: PF000053; laminin EGF_13.
 DR PFAM: PF000055; laminin Nterm; 1.
 DR PRINTS: PR00011; EGF_LAMININ.
 DR PROSITE: PS00022; EGF_1; 10.
 DR PROSITE: PS01186; EGF_2; 2.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; 12.
 DR Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 KW laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 FT SIGNAL 1 25
 FT CHAIN 26 1790 LAMININ BETA-1 CHAIN.
 FT DOMAIN 43 289 LAMININ N-TERMINAL (DOMAIN VI).
 FT DOMAIN 290 561 4.5 X LAMININ EGF-LIKE REPEATS (DOMAIN
 FT V).
 FT DOMAIN 290 356 LAMININ EGF-LIKE 1.
 FT DOMAIN 357 419 LAMININ EGF-LIKE 2.
 FT DOMAIN 420 479 LAMININ EGF-LIKE 3.
 FT DOMAIN 480 530 LAMININ EGF-LIKE 4.
 FT DOMAIN 531 561 LAMININ EGF-LIKE 5 (INCOMPLETE).
 FT DOMAIN 562 789 LAMININ DOMAIN IV.
 FT DOMAIN 791 1190 8 X LAMININ EGF-LIKE REPEATS (DOMAIN
 FT III).
 FT DOMAIN 791 838 LAMININ EGF-LIKE 6.
 FT DOMAIN 839 884 LAMININ EGF-LIKE 7.
 FT DOMAIN 885 934 LAMININ EGF-LIKE 8.
 FT DOMAIN 935 992 LAMININ EGF-LIKE 9.
 FT DOMAIN 993 1044 LAMININ EGF-LIKE 10.
 FT DOMAIN 1045 1095 LAMININ EGF-LIKE 11.
 FT DOMAIN 1096 1144 LAMININ EGF-LIKE 12.
 FT DOMAIN 1144 1190 LAMININ EGF-LIKE 13.
 FT DOMAIN 1191 1407 LAMININ II.
 FT DOMAIN 1408 1434 LAMININ ALPHA.
 FT DOMAIN 1435 1790 LAMININ I.
 FT DOMAIN 1257 1407 COILED COIL (POTENTIAL).
 FT DOMAIN 1455 1507 COILED COIL (POTENTIAL).
 FT DOMAIN 1542 1563 COILED COIL (POTENTIAL).
 FT DOMAIN 1610 1764 COILED COIL (POTENTIAL).
 FT DISULFID 290 299 BY SIMILARITY.
 FT DISULFID 292 320 BY SIMILARITY.
 FT DISULFID 322 331 BY SIMILARITY.
 FT DISULFID 334 354 BY SIMILARITY.
 FT DISULFID 357 366 BY SIMILARITY.
 FT DISULFID 387 396 BY SIMILARITY.
 FT DISULFID 399 417 BY SIMILARITY.
 FT DISULFID 420 433 BY SIMILARITY.
 FT DISULFID 422 448 BY SIMILARITY.
 FT DISULFID 450 459 BY SIMILARITY.
 FT DISULFID 462 477 BY SIMILARITY.
 FT DISULFID 480 493 BY SIMILARITY.
 FT DISULFID 482 500 BY SIMILARITY.
 FT DISULFID 502 511 BY SIMILARITY.
 FT DISULFID 514 528 BY SIMILARITY.
 FT DISULFID 791 803 BY SIMILARITY.

```

FT DISULFID 793 810 BY SIMILARITY.
FT DISULFID 812 821 BY SIMILARITY.
FT DISULFID 824 836 BY SIMILARITY.
FT DISULFID 839 851 BY SIMILARITY.
FT DISULFID 841 858 BY SIMILARITY.
FT DISULFID 860 869 BY SIMILARITY.
FT DISULFID 872 882 BY SIMILARITY.
FT DISULFID 885 894 BY SIMILARITY.
FT DISULFID 904 901 BY SIMILARITY.
FT DISULFID 904 913 BY SIMILARITY.
FT DISULFID 916 932 BY SIMILARITY.
FT DISULFID 935 951 BY SIMILARITY.
FT DISULFID 937 962 BY SIMILARITY.
FT DISULFID 964 973 BY SIMILARITY.
FT DISULFID 976 990 BY SIMILARITY.
FT DISULFID 993 1007 BY SIMILARITY.
FT DISULFID 995 1014 BY SIMILARITY.
FT DISULFID 1017 1026 BY SIMILARITY.
FT DISULFID 1029 1042 BY SIMILARITY.
FT DISULFID 1045 1059 BY SIMILARITY.
FT DISULFID 1047 1066 BY SIMILARITY.
FT DISULFID 1068 1077 BY SIMILARITY.
FT DISULFID 1080 1093 BY SIMILARITY.
FT DISULFID 1096 1108 BY SIMILARITY.
FT DISULFID 1098 1115 BY SIMILARITY.
FT DISULFID 1117 1126 BY SIMILARITY.
FT DISULFID 1129 1141 BY SIMILARITY.
FT DISULFID 1144 1156 BY SIMILARITY.
FT DISULFID 1146 1163 BY SIMILARITY.
FT DISULFID 1165 1174 BY SIMILARITY.
FT DISULFID 1177 1188 BY SIMILARITY.
FT DISULFID 1191 1191 INTERCHAIN (PROBABLE).
FT DISULFID 1194 1194 INTERCHAIN (PROBABLE).
FT DISULFID 1788 1788 INTERCHAIN (PROBABLE).
FT SITE 643 645 CELL ATTACHMENT SITE (POTENTIAL).
FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 203 203 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 593 593 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1053 1053 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1248 1248 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1303 1303 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1332 1332 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1343 1343 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 26.7% Score 47; DB 1; Length 1790;
Best Local Similarity 43.8% Pred. No. 1e+02;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Oy 12 CVPQINRGSVTNAG 27
    ||| | : | : | : |
Db 477 CEPCTCNPLGTLNNSG 492

RESULT 9
PRIL_PLAFK STANDARD; PRT; 452 AA.
ID PRIL_PLAFK
AC Q25998;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE DNA PRIMASE SMALL SUBUNIT (EC 2.7.7.-) (DNA PRIMASE 53 KDA SUBUNIT).
OS Plasmodium falciparum (isolate KI / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97078669; Pubmed=8918794;
RA Prasatkew S., Zijlstra N.M., Wilairat P., Overduin J.P.,
RT "Molecular cloning of a Plasmodium falciparum gene interrupted by 15
    introns encoding a functional primase 53 kDa subunit as demonstrated
    by expression in a baculovirus system."

```

```

RL Nucleic Acids Res. 24:3934-3941(1996).
CC -1- FUNCTION: DNA PRIMASE IS THE POLYMERASE THAT SYNTHESIZES SMALL
CC RNA PRIMERS FOR THE OKAZAKI FRAGMENTS MADE DURING DISCONTINUOUS
CC DNA REPLICATION (BY SIMILARITY).
CC -1- SUBUNIT: HETERODIMER OF A SMALL SUBUNIT AND A LARGE SUBUNIT.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC PRIMASE SMALL SUBUNIT
CC FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X99254; CAA67626.1; -
DR INTERPRO; IPR002755; -
DR PFAM; PF01896; DNA_primase-S; 1
KW Transferase; DNA replication; DNA-directed RNA polymerase; Primosome.
FT ACT_SITE 59 59 POTENTIAL.
FT ACT_SITE 130 130 POTENTIAL.
FT ACT_SITE 132 132 POTENTIAL.
FT METAL 142 142 POTENTIAL.
FT METAL 143 143 POTENTIAL.
FT METAL 149 149 POTENTIAL.
FT METAL 152 152 POTENTIAL.
SQ SEQUENCE 452 AA; 53489 MW; D8028925/445E8D6 CRC64;

Query Match 26.4% Score 46.5; DB 1; Length 452;
Best Local Similarity 31.2% Pred. No. 33;
Matches 10; Conservative 6; Mismatches 15; Indels 1; Gaps 1;

Oy 1 PPANNAARGICVPCQINRGSVTNAGDLATL 32
    ||| | : | : | : |
Db 361 PFCINHGTCGVCPIDKINNF-NPOSVPTL 391

RESULT 10
ENO_HOMGA STANDARD; PRT; 433 AA.
ID ENO_HOMGA
AC P56252;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE ENOLASE (EC 4.2.1.11) (2-PHOSHOGLYCERATE DEHYDRATASE) (2-PHOSPHO-D-
    GLYCERATE HYDRO-LYASE).
OS Homarus gammarus (European lobster) (Homarus vulgaris).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Nephropoidea; Nephropidae; Homarus.
RN [1]
RP SEQUENCE, AND X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RC TISSUE=MUSCLE;
RX MEDLINE=96038209; Pubmed=7547999;
RA Duquerry S., Camus C., Janin J.;
RT "X-ray structure and catalytic mechanism of lobster enolase."
RL Biochemistry 34:12513-12523(1995).
CC -1- CATALYTIC ACTIVITY: 2-PHOSPHO-D-GLYCERATE = PHOSPHOENOLPYRUVATE
    + H(2O).
CC -1- COFACTOR: MAGNESIUM IS REQUIRED FOR CATALYSIS AND FOR STABILIZING
    THE DIMER.
CC -1- PATHWAY: GLYCOLYSIS.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE ENOLASE FAMILY.
DR PDB; 1PDZ; 14-NOV-95.
DR INTERPRO; IPR000941; -
DR PFAM; PF00113; enolase; 1.
DR PRINTS; PR00148; ENOLASE.
DR PROSITE; PS00164; ENOLASE; 1.

```


CC	WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
CC	-1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
CC	DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
CC	TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
CC	COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
CC	THE BETA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ), LAMININ-
CC	2 (MEMOSIN), AND LAMININ-6 (K-LAMININ).
CC	-1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC	-1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
CC	COMPONENT).
CC	-1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC	WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC	-1- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
CC	-1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC	-1- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.
CC	-1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL, M61951, AAA59486.1, JOINED.
DR	EMBL, M58147, AAA59486.1, JOINED.
DR	EMBL, M61917, AAA59486.1, JOINED.
DR	EMBL, M61918, AAA59486.1, JOINED.
DR	EMBL, M61921, AAA59486.1, JOINED.
DR	EMBL, M61922, AAA59486.1, JOINED.
DR	EMBL, M61923, AAA59486.1, JOINED.
DR	EMBL, M61924, AAA59486.1, JOINED.
DR	EMBL, M61925, AAA59486.1, JOINED.
DR	EMBL, M61926, AAA59486.1, JOINED.
DR	EMBL, M61927, AAA59486.1, JOINED.
DR	EMBL, M61928, AAA59486.1, JOINED.
DR	EMBL, M61929, AAA59486.1, JOINED.
DR	EMBL, M61930, AAA59486.1, JOINED.
DR	EMBL, M61931, AAA59486.1, JOINED.
DR	EMBL, M61932, AAA59486.1, JOINED.
DR	EMBL, M61933, AAA59486.1, JOINED.
DR	EMBL, M61934, AAA59486.1, JOINED.
DR	EMBL, M61935, AAA59486.1, JOINED.
DR	EMBL, M61936, AAA59486.1, JOINED.
DR	EMBL, M61938, AAA59486.1, JOINED.
DR	EMBL, M61939, AAA59486.1, JOINED.
DR	EMBL, M61940, AAA59486.1, JOINED.
DR	EMBL, M61941, AAA59486.1, JOINED.
DR	EMBL, M61942, AAA59486.1, JOINED.
DR	EMBL, M61943, AAA59486.1, JOINED.
DR	EMBL, M61944, AAA59486.1, JOINED.
DR	EMBL, M61945, AAA59486.1, JOINED.
DR	EMBL, M61946, AAA59486.1, JOINED.
DR	EMBL, M61947, AAA59486.1, JOINED.
DR	EMBL, M61948, AAA59486.1, JOINED.
DR	EMBL, M61949, AAA59486.1, JOINED.
DR	EMBL, M61950, AAA59486.1, JOINED.
DR	EMBL, M55370, AAA59485.1, JOINED.
DR	EMBL, M55378, AAA59485.1, JOINED.
DR	EMBL, M53565, AAA59485.1, JOINED.
DR	EMBL, M53571, AAA59485.1, JOINED.
DR	EMBL, M55372, AAA59485.1, JOINED.
DR	EMBL, M55373, AAA59485.1, JOINED.
DR	EMBL, M55374, AAA59485.1, JOINED.
DR	EMBL, M53575, AAA59485.1, JOINED.
DR	EMBL, M53576, AAA59485.1, JOINED.
DR	EMBL, M55344, AAA59485.1, JOINED.
DR	EMBL, M55345, AAA59485.1, JOINED.
DR	EMBL, M55346, AAA59485.1, JOINED.
DR	EMBL, M55347, AAA59485.1, JOINED.
DR	EMBL, M55348, AAA59485.1, JOINED.
DR	EMBL, M55349, AAA59485.1, JOINED.

DR	EMBL	M55350	AAAS9485.1	JOINED
DR	EMBL	M55351	AAAS9485.1	JOINED
DR	EMBL	M55352	AAAS9485.1	JOINED
DR	EMBL	M55353	AAAS9485.1	JOINED
DR	EMBL	M55354	AAAS9485.1	JOINED
DR	EMBL	M55355	AAAS9485.1	JOINED
DR	EMBL	M55356	AAAS9485.1	JOINED
DR	EMBL	M55357	AAAS9485.1	JOINED
DR	EMBL	M55358	AAAS9485.1	JOINED
DR	EMBL	M55359	AAAS9485.1	JOINED
DR	EMBL	M55360	AAAS9485.1	JOINED
DR	EMBL	M55361	AAAS9485.1	JOINED
DR	EMBL	M55362	AAAS9485.1	JOINED
DR	EMBL	M55363	AAAS9485.1	JOINED
DR	EMBL	M55364	AAAS9485.1	JOINED
DR	EMBL	M55365	AAAS9485.1	JOINED
DR	EMBL	M55367	AAAS9485.1	JOINED
DR	EMBL	M55368	AAAS9485.1	JOINED
DR	EMBL	M55369	AAAS9485.1	JOINED
DR	EMBL	M61916	AAAS9482.1	..
DR	EMBL	M20206	AAAS9487.1	..
DR	PIR	S13547	MMHUB1	
DR	HSSP	P01543	1BHP	
DR	MIM	150240	-	
DR	INTERPRO	IPR000561	-	
DR	INTERPRO	IPR001886	-	
DR	INTERPRO	IPR002049	-	
DR	PFAM	PF00053	Laminin_EGF_13	
DR	PFAM	PF00055	Laminin_Nterm_1	
DR	PRINTS	PR00011	EGFLAMIN	
DR	PROSITE	PS00022	EGF_1; 9	
DR	PROSITE	PS01166	EGF_2; 2	
DR	PROSITE	PS01246	LAMININ_Type_EGF_11	
DR	KW	Glycoprotein	Basement membrane; Cell adhesion; Repeat; Signal	
KW	Laminin	EGF-like domain		
FT	SIGNAL	1	21	
FT	CHAIN	22	1786	
FT	DOMAIN	22	270	
FT	DOMAIN	271	540	
FT	DOMAIN	271	334	LAMININ EGF-LIKE 1.
FT	DOMAIN	335	397	LAMININ EGF-LIKE 2.
FT	DOMAIN	398	457	LAMININ EGF-LIKE 3.
FT	DOMAIN	458	509	LAMININ EGF-LIKE 4.
FT	DOMAIN	510	540	LAMININ EGF-LIKE 5 (INCOMPLETE).
FT	DOMAIN	541	771	LAMININ DOMAIN IV.
FT	DOMAIN	772	1178	LAMININ EGF-LIKE REPEATS (DOMAIN III).
FT	DOMAIN	773	820	LAMININ EGF-LIKE 6.
FT	DOMAIN	821	866	LAMININ EGF-LIKE 7.
FT	DOMAIN	867	916	LAMININ EGF-LIKE 8.
FT	DOMAIN	917	975	LAMININ EGF-LIKE 9.
FT	DOMAIN	976	1027	LAMININ EGF-LIKE 10.
FT	DOMAIN	1028	1083	LAMININ EGF-LIKE 11.
FT	DOMAIN	1084	1131	LAMININ EGF-LIKE 12.
FT	DOMAIN	1132	1178	LAMININ EGF-LIKE 13.
FT	DOMAIN	1179	1397	DOMAIN II.
FT	DOMAIN	1398	1430	DOMAIN ALPHA.
FT	DOMAIN	1431	1786	DOMAIN I.
FT	DOMAIN	1216	1315	COILED COIL (POTENTIAL).
FT	DOMAIN	1353	1388	COILED COIL (POTENTIAL).
FT	DOMAIN	1442	1781	COILED COIL (POTENTIAL).
FT	DISULFID	271	280	BY SIMILARITY.
FT	DISULFID	273	298	BY SIMILARITY.
FT	DISULFID	300	309	BY SIMILARITY.
FT	DISULFID	312	332	BY SIMILARITY.
FT	DISULFID	335	344	BY SIMILARITY.
FT	DISULFID	337	362	BY SIMILARITY.
FT	DISULFID	365	374	BY SIMILARITY.
FT	DISULFID	377	395	BY SIMILARITY.
FT	DISULFID	398	411	BY SIMILARITY.
FT	DISULFID	400	426	BY SIMILARITY.
FT	DISULFID	428	437	BY SIMILARITY.
FT	DISULFID	440	455	BY SIMILARITY.

```

FT DISULFID 458 472 BY SIMILARITY.
RL DISULFID 460 479 BY SIMILARITY.
FT DISULFID 481 490 BY SIMILARITY.
FT DISULFID 493 507 BY SIMILARITY.
FT DISULFID 773 785 BY SIMILARITY.
FT DISULFID 775 792 BY SIMILARITY.
FT DISULFID 794 803 BY SIMILARITY.
FT DISULFID 806 818 BY SIMILARITY.
FT DISULFID 821 833 BY SIMILARITY.
FT DISULFID 823 840 BY SIMILARITY.
FT DISULFID 842 851 BY SIMILARITY.
FT DISULFID 854 864 BY SIMILARITY.
FT DISULFID 867 876 BY SIMILARITY.
FT DISULFID 869 883 BY SIMILARITY.
FT DISULFID 886 895 BY SIMILARITY.
FT DISULFID 898 914 BY SIMILARITY.
FT DISULFID 917 933 BY SIMILARITY.
FT DISULFID 919 944 BY SIMILARITY.
FT DISULFID 946 955 BY SIMILARITY.

```

Query Match 26.1%; Score 46; DB 1; Length 1786;
 Best Local Similarity 42.1%; Pred. No. 1.4e+02;
 Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

```

OY 3 AANNAAGICVPCQINRYG 21
Db 353 ATGNVSGVGCDDCOHNTWG 371

```

```

RESULT 13
ENOG_HUMAN STANDARD; PRT; 433 AA.
AC P09104;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE GAMMA ENOLASE (PC 4.2.1.11) (2-PHOSPHO-D-GLYCERATE HYDRO-LYASE)
DE (NEURAL ENOLASE) (NSE).
GN ENO2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=RETINA;
RX MEDLINE=89091176; Pubmed=3208766;
RA McAleese S.M., Dunbar B., Forcheggill J., Hinks L., Day I.N.M.;
RT "Complete amino acid sequence of the neuron-specific gamma isozyme
RT of enolase (NSE) from human brain and comparison with the
RT non-neuronal alpha form (NNE).";
RT Eur. J. Biochem. 178:413-417(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9006764; Pubmed=2792767;
RA Oliva D., Barba G., Barbieri G., Giallongo A., Feo S.;
RT "Cloning, expression and sequence homologies of cDNA for human gamma
RT enolase.";
RT Gene 79:355-360(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=88259288; Pubmed=3385803;
RA van Obberghen E., Kamholz J., Bishop J.G. III, Zomzely-Neurath C.,
RA Lazzarini R.A.;
RT "Human gamma enolase: isolation of a cDNA clone and expression in
RT normal and tumor tissues of human origin.";
RT J. Neurosci. Res. 19:450-456(1988).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=HEMATOPOIETIC;
RX MEDLINE=91257823; Pubmed=2045099;
RA Oliva D., Cali L., Feo S., Giallongo A.;
RT "Complete structure of the human gene encoding neuron-specific

```

```

RT enolase.";
RL Genomics 10:157-165(1991).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=96303695; Pubmed=8723724;
RA Ansari-Lari M.A., Mazy D.M., Lu J., Lu F., Lilley C.E.,
RA Spanos S., Malley T., Gibbs R.A.;
RT "A gene-rich cluster between the CD4 and triosephosphate isomerase
RT genes at human chromosome 12p13.";
RL Genome Res. 6:314-326(1996).
RN [6]
RP SEQUENCE OF 424-433 FROM N.A.
RX MEDLINE=88005129; Pubmed=3653393;
RA Day I.N.M., Allsopp M.T.E.P., Moore D.C.M., Thompson R.J.;
RT "Sequence conservation in the 3'-untranslated regions of neurone-
RT specific enolase, lymphokine and protooncogene mRNAs.";
RL FEBS Lett. 222:139-143(1987).
CC -1- CATALYTIC ACTIVITY: 2-PHOSPHO-D-GLYCERATE = PHOSPHOENOLPYRUVATE
CC + H(2O).
CC -1- COFACTOR: MAGNESIUM IS REQUIRED FOR CATALYSIS AND FOR STABILIZING
CC THE DIMER.
CC -1- PATHWAY: GLYCOLYSIS.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- TISSUE SPECIFICITY: THREE CLASSES OF ENOLASE ISOENZYMES HAVE BEEN
CC IDENTIFIED IN MAMMALIAN TISSUES. ISOENZYME ALPHA IS PRESENT IN
CC MOST TISSUES, BETA IS LOCALIZED IN MUSCLE TISSUE, AND GAMMA IS
CC FOUND ONLY IN NERVOUS TISSUE.
CC -1- SIMILARITY: BELONGS TO THE ENOLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL; X13120; CA831512.1; -
DR EMBL; M22349; AAB59554.1; -
DR EMBL; M36768; AAA52388.1; ALT_INIT.
DR EMBL; X51956; CA836215.1; -
DR EMBL; U47924; CAB35670.1; -
DR EMBL; X14327; CA832505.1; -
DR PIR; J00060; NOHUG.
DR PIR; S02077; S02077.
DR PIR; S02616; S02616.
DR HSSP; P56252; 1PDY.
DR MIM; 131360; -
DR INTERPRO; IPR000941; -
DR PFM; PF00113; enolase; 1.
DR PRINTS; PR00148; ENOLASE.
DR PROSITE; PS00164; ENOLASE; 1.
KW Lyase; Glycolysis; Magnesium; Multigene family; Polymorphism.
FT INT_MET 0
FT ACT_SITE 157 157
FT METAL 244 244
FT METAL 292 292
FT METAL 317 317
FT VARIANT 263 263
FT P->A.
FT /FTID=VAR_002354.
FT T->A.
FT /FTID=VAR_002355.
FT O->E (IN REF. 2, 4 AND 5).
FT AK->G (IN REF. 3).
FT E->N (IN REF. 3).
FT M->I (IN REF. 2, 3, 4 AND 5).
SO SEQUENCE 433 AA; 47154 MW; 0F01A8A5FF9152A1 CRC64;

```

Query Match 25.6%; Score 45; DB 1; Length 433;
 Best Local Similarity 40.9%; Pred. No. 51;
 Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

OY 12 CVPQINRVGSVTNAGDLATLA 33
 Db 338 CILKVNQIGSVTEAIOACKLA 359

RESULT 14
 ENOG_MOUSE STANDARD: PRT: 433 AA.

AC P17183;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE GAMMA ENOLASE (EC 4.2.1.11) (2-PHOSPHO-D-GLYCERATE HYDRO-LYASE)
 DE (NEURAL ENOLASE) (NSE)
 GN ENO2 OR ENO-2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE=90301487; PubMed=2362815;
 RA Kaghad M., Dumont X., Chalon P., Lelias J.M., Lamande N., Lucas M.,
 RA Lazar M., Caput D.;
 RT "Nucleotide sequences of cDNAs alpha and gamma enolase mRNAs from
 RT mouse brain."
 RL Nucleic Acids Res. 18:3638-3638(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Ansaari-Lari M.A., Oel'tjen J.C., Schwartz S., Zhang Z., Muzny D.M.,
 RA Lu J., Gorrell J.H., Chinault A.C., Belmont J.W., Miller W.,
 RA Gdbbs R.A.;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: 2-PHOSPHO-D-GLYCERATE = PHOSPHOENOLPYRUVATE
 CC + H(2)O.
 CC -1- COFACTOR: MAGNESIUM IS REQUIRED FOR CATALYSIS AND FOR STABILIZING
 CC THE DIMER.
 CC -1- PATHWAY: GLYCOLYSIS.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- TISSUE SPECIFICITY: THREE CLASSES OF ENOLASE ISOENZYMES HAVE BEEN
 CC IDENTIFIED IN MAMMALIAN TISSUES. ISOENZYME ALPHA IS PRESENT IN
 CC MOST TISSUES, BETA IS LOCALIZED IN MUSCLE TISSUE, AND GAMMA IS
 CC FOUND ONLY IN NERVOUS TISSUE.
 CC -1- SIMILARITY: BELONGS TO THE ENOLASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X52380; CAA36606.1; -
 DR EMBL: AC002397; AAC36002.1; -
 DR PIR: S10247; S10247.
 DR HSSP: P56252; IPDY.
 DR MGD: MGI:95394; ENO2.
 DR INTERPRO: IPR000941; -
 DR PFM: PFO0113; enolase; 1.
 DR PRINTS: PR00148; ENOLASE.
 DR PROSITE: PS00164; ENOLASE; 1.
 KW Lyase; Glycolysis; Magnesium; Multigene family.
 FT INIT_MET 0
 FT ACT_SITE 157
 FT METAL 244
 FT METAL 292
 FT METAL 317
 SQ SEQUENCE 433 AA; 47165 MW; B7281132E637D43E CRC64;

Query Match 25.6%; Score 45; DB 1; Length 433;
 Best Local Similarity 40.9%; Pred. No. 51;
 Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

OY 12 CVPQINRVGSVTNAGDLATLA 33
 Db 338 CILKVNQIGSVTEAIOACKLA 359

RESULT 15

ENOG_RAT STANDARD: PRT: 433 AA.

AC P07323;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE GAMMA ENOLASE (EC 4.2.1.11) (2-PHOSPHO-D-GLYCERATE HYDRO-LYASE)
 DE (NEURAL ENOLASE) (NSE).
 GN ENO2 OR ENO-2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86042683; PubMed=2865729;
 RA Sakimura K., Kushiya E., Obinata M., Odani S., Takahashi Y.;
 RT "Molecular cloning and the nucleotide sequence of cDNA for neuron-
 RT specific enolase messenger RNA of rat brain."
 RL Proc. Natl. Acad. Sci. U.S.A. 82:7453-7457(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY;
 RX MEDLINE=86308095; PubMed=3746946;
 RA Forss-Petter S., Danielson P., Sutcliffe J.G.;
 RT "Neuron-specific enolase: complete structure of rat mRNA, multiple
 RT transcriptional start sites, and evidence suggesting post-
 RT transcriptional control."
 RL J. Neurosci. Res. 16:141-156(1986).
 RN [3]
 RP SEQUENCE OF 1-27 FROM N.A.
 RX MEDLINE=88152493; PubMed=2450052;
 RA Sakimura K., Kushiya E., Takahashi Y., Suzuki Y.;
 RT "The structure and expression of neuron-specific enolase gene."
 RL Gene 60:103-113(1987).
 CC -1- CATALYTIC ACTIVITY: 2-PHOSPHO-D-GLYCERATE = PHOSPHOENOLPYRUVATE
 CC + H(2)O.
 CC -1- COFACTOR: MAGNESIUM IS REQUIRED FOR CATALYSIS AND FOR STABILIZING
 CC THE DIMER.
 CC -1- PATHWAY: GLYCOLYSIS.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- TISSUE SPECIFICITY: THREE CLASSES OF ENOLASE ISOENZYMES HAVE BEEN
 CC IDENTIFIED IN MAMMALIAN TISSUES. ISOENZYME ALPHA IS PRESENT IN
 CC MOST TISSUES, BETA IS LOCALIZED IN MUSCLE TISSUE, AND GAMMA IS
 CC FOUND ONLY IN NERVOUS TISSUE.
 CC -1- SIMILARITY: BELONGS TO THE ENOLASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M11931; AAA41119.1; -
 DR EMBL: M22770; AAA41725.1; -
 DR EMBL: AF019973; AAB72088.1; -
 DR EMBL: X07727; CAA30556.1; -
 DR EMBL: X07728; CAA30556.1; JOINED.
 DR EMBL: X07729; CAA30556.1; JOINED.
 DR PIR: A24742; A24742.
 DR HSSP: P56252; IPDY.

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:49:33 ; Search time 83.05 Seconds
(without alignments)
8.235 Million cell updates/sec

Title: US-09-196-161d-5
Perfect score: 106
Sequence: 1 CPTGALDGVTVDFRASA 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues
Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_36:*

- 1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47.5	44.8	61	19	W76655
2	46	43.4	73	21	Y52238
3	46	43.4	362	18	W37361
4	46	43.4	463	21	Y52216
5	44	41.5	355	20	Y33689
6	44	41.5	424	17	W04321
7	43	40.6	461	20	W89558
8	42	39.6	19	13	R22003
9	42	39.6	19	14	R41802
10	42	39.6	20	14	R41798
11	42	39.6	70	19	W76685
12	42	39.6	114	14	R41799

13	42	39.6	114	16	R84184	Human derived thro
14	42	39.6	114	18	W01599	Thrombomodulin TMD
15	42	39.6	115	11	R07074	Anticoagulant and
16	42	39.6	115	13	R22004	Coagulation and pl
17	42	39.6	115	17	R94610	Human-recombinant
18	42	39.6	116	11	R06831	Thrombomodulin ana
19	42	39.6	118	14	R06833	Thrombomodulin ana
20	42	39.6	118	14	R34679	Thrombomodulin ana
21	42	39.6	118	19	R43679	E456 part of EGF-1
22	42	39.6	130	21	W41367	Thrombomodulin for
23	42	39.6	130	21	Y69528	Human thrombomodul
24	42	39.6	131	13	R22015	Human thrombomodul
25	42	39.6	132	13	R22014	Human thrombomodul
26	42	39.6	132	13	R22019	Human thrombomodul
27	42	39.6	132	21	Y83936	Human thrombomodul
28	42	39.6	133	13	R22020	Human thrombomodul
29	42	39.6	133	13	R22022	Human thrombomodul
30	42	39.6	134	13	R22021	Human thrombomodul
31	42	39.6	166	14	R34682	Human thrombomodul
32	42	39.6	182	21	Y87519	tPA signal peptide
33	42	39.6	229	9	P82071	Bacillus thuringie
34	42	39.6	240	11	R06827	Proel1743-2-9-8 in
35	42	39.6	242	11	R06828	Thrombomodulin ana
36	42	39.6	249	11	R06826	Thrombomodulin ana
37	42	39.6	253	11	R06825	Thrombomodulin ana
38	42	39.6	253	11	R04241	Thrombin-binding p
39	42	39.6	253	17	R88514	Thrombomodulin-lik
40	42	39.6	254	21	Y83937	Human thrombomodul
41	42	39.6	254	21	Y83938	Human thrombomodul
42	42	39.6	275	11	R06832	Human thrombomodul
43	42	39.6	275	11	R94609	Human recombinant
44	42	39.6	277	11	R06834	Thrombomodulin ana
45	42	39.6	331	9	P82847	Sequence of new pe

ALIGNMENTS

RESULT 1	
W76655	W76655 standard; Protein; 61 AA.
XX	
AC	W76655;
XX	
DT	05-JAN-1999 (first entry)
XX	
DE	Elapidae modified dendroaspin protein fragment DEN-TM2.
XX	
KW	Dendroaspin: snake venom; clotting cascade; anticoagulant; platelet;
KW	integrin binding; injury; blood; cell migration; thrombosis; inhibitor;
KW	proliferation; signal transduction; regulator; coagulation; treatment;
KW	prophylactic; artery; vein; wall thickening; myocardial infarction;
KW	retinal neovascularisation; dysregulated apoptosis; tumorigenesis;
KW	leukocyte recruitment; immune system; tissue fibrosis.
XX	
OS	Elapidae.
OS	Synthetic.
XX	
FN	W09842834-A1.
XX	
PD	01-OCT-1998.
XX	
PF	20-MAR-1998; 98WO-GB00848.
XX	
PR	20-MAR-1997; 97GB-0005787.
XX	
PA	(THRO-) THROMBOSIS RES INST.
XX	
PI	Authi K, Kakkar V, Lu X, Scully MF;
XX	
DR	WPI: 1998-542278/46.
XX	

New hybrid dendroaspin polypeptide(s) - used for treating, e.g.

thrombosis, myocardial infarction, dysregulated apoptosis, abnormal cell migration and immune system activation

Claim 6; Fig 3A; 59pp; English.

W6645-W6688 represent modified dendroaspin protein fragments isolated from snake venom. When dendroaspin is modified to incorporate further functional amino acid sequence, e.g. active portions or motifs of agonists, antagonists or inhibitors of factors in the clotting cascade, the resulting molecules are particularly useful as anticoagulants. The molecules have an integrin binding activity which when administered in vivo results in the binding of the molecules to platelets thereby inhibiting the aggregation of the platelets at sites of injury. Non-wild type dendroaspin domains provide secondary, optionally further functionality, e.g. antithrombotic action, inhibiting cell migration and proliferation and regulating signal transduction. Such variants have bi- or multifunctional activities against blood coagulation, particularly thrombus formation and arterial/venous wall thickening at the sites of injury. The variants may have activities against leukocyte recruitment, immune system activation, tissue fibrosis and tumourigenesis. The polypeptides can be used for the treatment or prophylaxis of a disease associated with thrombosis, e.g. myocardial infarction, retinal neovascularisation, endothelial injury, dysregulated apoptosis, abnormal cell migration, leukocyte recruitment, immune system activation, tissue fibrosis or tumorigenesis.

Sequence 61 AA:

Query Match 44.88; Score 47.5; DB 19; Length 61;
 Best Local Similarity 60.08; Pred. No. 0.5;
 Matches 9; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

1 CPTGTALDDG-VTDV 14
 ||| ||||| :|||
 Db 22 cpegrliddgfildl 36

RESULT 2
 Y52238
 ID Y52238 standard; protein; 73 AA.
 XX Y52238;
 AC
 XX 09-FEB-2000 (first entry)
 DT
 XX
 DE A. thaliana yjld protein fragment containing conserved motif 1.
 DE
 XX
 KW yjld protein; essential; Gram positive; Gram negative; conserved;
 KW motif; identification; antagonist; antibacterial; antibiotic;
 KW broad spectrum; treatment; infection; resistance; drug target.
 OS
 XX Arabidopsis thaliana.
 OS
 XX
 FT Key Location/Qualifiers
 FT Region 10..61
 FT /note= "yjld conserved motif 1"
 XX
 XX WO954470-A2.
 XX
 XX PD 28-OCT-1999.
 XX
 XX PF 20-APR-1999; 99WO-EP02635.
 XX
 XX PR 22-APR-1998; 98GB-0008423.
 XX
 XX PA (GLAX) GLAXO GROUP LTD.
 XX
 XX PI Arigoni F, Edgerton MD, Loferer H, Peitsch MC;
 XX WPI; 2000-013253/01.
 XX DR
 XX Novel bacterial polypeptides used to identify broad spectrum

antibiotics

Claim 3; Fig 2a; 55pp; English.

Proteins Y5223-Y5238 represent fragments of yjld proteins containing a conserved motif (motif 1). yjld proteins are essential for the survival of both Gram negative and Gram positive bacteria, although no function has as yet been ascribed to these proteins. The yjld proteins, fragments of yjld proteins (for example, fragments encompassing one or more conserved yjld motifs such as Y5218-Y5284) and nucleotides encoding them can be used to identify antagonists and broad spectrum antibacterial compounds. These antagonists and compounds can be used to treat a wide range of bacterial infections. New antibiotics are urgently needed, as serious bacterial infections and antibiotic resistant strains are becoming increasingly prevalent. The proteins of the invention are essential proteins for bacterial viability, and represent new targets for antibiotics.

Sequence 73 AA:

Query Match 43.48; Score 46; DB 21; Length 73;
 Best Local Similarity 43.88; Pred. No. 1.1;
 Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

4 GTALDDGVTDVDRSA 19
 ||:|:|:|:|:|
 Db 38 gtlvdaigaefdkta 53

RESULT 3
 W37361
 ID W37361 standard; protein; 362 AA.
 XX W37361;
 AC
 XX 27-APR-1998 (first entry)
 DT
 XX
 DE PsbI (Wbpi) protein involved in O-antigen synthesis and assembly.
 DE
 XX
 KW O antigen; PsbI; Wbpi; lipopolysaccharide; infection; diagnosis;
 KW antibody; UDP-N-acetylmannosamine.
 KW
 XX Pseudomonas aeruginosa PAOI.
 OS
 XX
 XX WO9741234-A2.
 XX
 XX PD 06-NOV-1997.
 XX
 XX PF 30-APR-1997; 97WO-CA00295.
 XX
 XX PR 27-FEB-1997; 97US-0039473.
 XX
 XX PR 30-APR-1996; 96US-0016510.
 XX
 XX PA (UYGU-) UNIV GUELPH.
 XX
 XX PI Burrows L, Charter D, De Kievit T, Lam JS;
 XX WPI; 1997-549736/50.
 XX DR N-PSDB; T97221.
 XX
 XX PT Pseudomonas aeruginosa B-band lipopolysaccharide gene cluster -
 XX used for diagnosis of P. aeruginosa infection
 XX
 XX PS Claim 8; Page 120-121; 195pp; English.
 XX
 XX This sequence comprises PsbH (WbPH), a Pseudomonas aeruginosa PAOI
 XX protein that converts UDP-N-acetylglucosamine to
 XX UDP-N-acetylmannosamine. Wzz (Rol1), PsbA (WbPA), PsbB (WbPB), PsbC
 XX (WbPC), PsbD (WbPD), PsbE (WbPE), Rfc (Wzy), PsbF (WbPF), PsbG
 XX (WbPG), PsbH (WbPH), PsbI (WbPI), PsbJ (WbPJ), PsbK (WbPK), PsbM
 XX (WbPM) and PsbN (WbPN) (see W37349-56 and W37357-65, respectively)
 XX are claimed. They are involved in the synthesis and assembly of

CC B-band lipopolysaccharide (i.e. O-antigen). Also claimed are: UVRB
 CC (W37366) involved in ultraviolet repair and Hish and Hish
 CC (W37357-58) involved in histidine synthesis. All these proteins
 CC are encoded by the B-band gene cluster (see T97221) of P.
 CC aeruginosa PA01. Purified proteins can be obtained from
 CC transformed host cells and used to raise monoclonal or polyclonal
 CC antibodies. Such antibodies specifically recognise the B-band
 CC lipopolysaccharide and can be used in a claimed method for
 CC detecting P. aeruginosa in a sample, i.e. to diagnose infection.
 CC
 SQ Sequence 362 AA;

Query Match 43.4%; Score 46; DB 18; Length 362;
 Best Local Similarity 45.0%; Pred. No. 7.2;
 Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 CPTGALDDGVTFDRSA 20
 ||| |::: |::: |
 Db 146 cplvaiddlknegierkaa 165

RESULT 4

ID Y52216 standard; protein; 463 AA.

AC Y52216;

DT 09-FEB-2000 (first entry)

DE Arabidopsis thaliana y5jd protein homologue.

KW y5jd protein; essential; Gram positive; Gram negative; conserved;
 KW motif; identification; antagonist; antibacterial; antibiotic;
 KW broad spectrum; treatment; infection; resistance; drug target.

OS Arabidopsis thaliana.

PN Key Location/Qualifiers

FT Region 86..96 /note= "y5jd conserved motif 3"

FT Region 111..131 /note= "y5jd conserved motif 4"

FT Region 152..198 /note= "y5jd conserved motif 2"

FT Region 208..259 /note= "y5jd conserved motif 1"

PN WO9954470-A2.

PD 28-OCT-1999.

PF 20-APR-1999; 99WO-EP02635.

PR 22-APR-1998; 98GB-0008423.

PA (GLAXO) GLAXO GROUP LTD.

PI Arigoni F, Edgerton MD, Loferer H, Reitsch MC;

DR WPI: 2000-013253/01.

PT Novel bacterial polypeptides used to identify broad spectrum

PS Claim 1; Fig 1; 55pp; English.

CC Proteins Y52202-Y52217 encompass a novel family of proteins
 CC designated the y5jd family, after the name given to the Escherichia
 CC coli family member. These proteins are essential for the survival of
 CC both Gram negative and Gram positive bacteria, although no function has
 CC as yet been ascribed to these proteins. The y5jd proteins, fragments of
 CC y5jd proteins (for example, fragments encompassing one or more

CC conserved y5jd motifs such as Y52218-Y52284) and nucleotides encoding
 CC them can be used to identify antagonists and broad spectrum antibacterial
 CC compounds. These antagonists and compounds can be used to treat a
 CC wide range of bacterial infections. New antibiotics are urgently needed,
 CC as serious bacterial infections and antibiotic resistant strains are
 CC becoming increasingly prevalent. The proteins of the invention are
 CC essential proteins for bacterial viability, and represent new targets
 CC for antibiotics.
 CC
 SQ Sequence 463 AA;

Query Match 43.4%; Score 46; DB 21; Length 463;
 Best Local Similarity 43.8%; Pred. No. 9.6;
 Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 4 GTALDDGVTFDRSA 19
 ||| |::: |::: |
 Db 235 gltvdaalgaardkia 250

RESULT 5

ID Y33689 standard; Protein; 355 AA.

AC Y33689;

DT 12-JAN-2000 (first entry)

DE Maize 33kDa cysteine proteinase Mirl protein fragment.

KW Cysteine proteinase; Mirl; maize; sterility; male plant; anther; cotton;
 KW insect resistance; lepidoptera; sugar cane; sugar beet; seed.

OS Zea mays.

PN US5977440-A.

PD 02-NOV-1999.

PF 30-APR-1996; 96US-0641314.

PR 30-APR-1996; 96US-0641314.

PA (LUTHR) LUTHE D S.

PA (WILL) WILLIAMS W P.

PA (JIANG) JIANG B.

PA (PECH) PECHAN T.

PI Williams WP, Luthe DS, Jiang B, Pechan T;

DR WPI: 1999-619723/53.

DR N-PSDB: 223708.

PT CDNA encoding maize cysteine protease useful for imparting insect

PS resistance or male sterility to plants, especially maize

PS Disclosure: Fig 1A-C; 12pp; English.

CC This invention describes a novel cDNA, called Mirl, encoding a maize
 CC 33 kD cysteine proteinase. An expression vector containing the entire
 CC Mirl sequence linked to one or more control elements can be used to
 CC impart sterility to a male plant by a method comprising transforming an
 CC anther of the plant with the vector and expressing the 33 kD cysteine
 CC proteinase in the anther. A baculovirus transformed with such a vector
 CC can be used to impart insect resistance to a plant by a method comprising
 CC infecting the plant with the transformed baculovirus. A vector containing
 CC a DNA molecule consisting of nucleotides 987-1250 of the Mirl sequence
 CC linked to a cDNA molecule encoding a cysteine proteinase and one or more
 CC control elements can be used to impart insect resistance to a plant
 CC susceptible to lepidopteran feeding, preferably a cotton, sugar cane,
 CC sugar beet or especially maize plant, by a method comprising introducing
 CC the vector into a plant seed and germinating the seed into a plant. This

CC sequence represents the maize M1r1 protein described in the method of the
CC invention.
XX
SQ Sequence 355 AA;

Query Match 41.5%; Score 44; DB 20; Length 355;
Best Local Similarity 69.2%; Pred. No. 15;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 PTGTALDDGVTDV 14
| | | | | | | | | |
Db 267 pcgtsldhgvclav 279

RESULT 6
W04321
ID W04321 standard; Protein; 424 AA.

AC W04321;

DT 17-JAN-1997 (first entry)

DE Ancylostoma secreted protein ASP-1 (pro-form).

KW Ancylostoma secreted protein; ASP-1; hookworm; vaccine.

OS Ancylostoma caninum.

FH Key Location/Qualifiers

FT Peptide 1..18

FT /label= Sig-peptide

FT Protein 19..424

FT /label= Mat_protein

PN W09632479-A1.

PD 17-OCT-1996.

PF 10-APR-1996; 96WO-US04821.

PR 10-APR-1995; 95US-0419414.

PA (UYVA) UNIV YALE.

PI Hawdon JM, Hotez PJ, Jones BF;

DR WPI: 1996-477130/47.

DR N-PSDB; T38466.

XX Ancylostoma caninum secreted protein - useful as antigen for

PT hookworm vaccine prodn.

PS Claim 2; Page 42-43; 66pp; English.

XX Ancylostoma secreted protein ASP-1 is secreted by canine hookworm

CC larvae as they change from the free-living stage to the parasitic

CC stage. The amino acid sequence of the ASP-1 pro-form (W04321) was

CC derivd. from a cDNA clone (T38466) obtd. from an Ancylostoma caninum

CC L3 larvae cDNA library. ASP-1 represents a family of proteins (see

CC also W04322-23) that are highly immunogenic in experimental

CC animals. Recombinant ASP proteins can be produced in a variety of

CC hosts. They can be used in vaccines for hookworm, in the diagnosis

CC of hookworm infection, or to raise antibodies.

Db 227 cpsnlgmldsvrldf 241
| | : | : | | | | |

RESULT 7

W89558
ID W89558 standard; Protein; 461 AA.

AC W89558;

DT 16-MAR-1999 (first entry)

DE Triticum sp. cysteine proteinase #3.

KW Triticum; wheat seed; cysteine proteinase; gluten; baking.

OS Triticum sp.

FH Key Location/Qualifiers

FT Misc-difference 290 /note= "encoded by CAT"

PN -JP10327886-A.

PD 15-DEC-1998.

PF 27-MAR-1998; 98JP-0098140.

PR 31-MAR-1997; 97JP-0114946.

PA (SHOS) SHOWA SANGYO CO.

DR WPI: 1999-109255/10.

DR N-PSDB; V82458.

PT New DNA coding cysteine proteinase originating from wheat seed -

FT useful for improving gluten for use in bakery process

PS Claim 3; Page 18-20; 29pp; Japanese.

CC The present sequence represents a cysteine proteinase isolated from

CC wheat seed (Triticum sp.). The cysteine proteinase is useful for

CC improving gluten for use in the bakery process.

XX Sequence 461 AA;

SQ

Query Match 40.6%; Score 43; DB 20; Length 461;
Best Local Similarity 81.8%; Pred. No. 30;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 4 GTALDDGVTDV 14
| | | | | | | | | |
Db 285 gtaiddgyvaav 295

RESULT 8
R22003
ID R22003 standard; Protein; 19 AA.

AC R22003;

DT 03-JUL-1992 (first entry)

DE Polypeptide C based on amino acids 406-444 of thrombomodulin.

KW Thrombin binding site; blood clotting.

OS Synthetic.

PN EP474273-A.

PD 11-MAR-1992.

```

XX 05-AUG-1991; 91EP-0202009.
PF 03-AUG-1990; 90JP-0204978.
PR (ASAH ) ASAH KASEI KOGYO.
PA
XX Zushi M, Gomi K, Yamamoto S, Suzuki K, Matsuda A:
XX WPI; 1992-081820/11.
DR
XX New polypeptide inhibits blood coagulation and platelet
PT aggregation - promotes thrombin catalysed protein C activation
PT for treating myocardial infarction, thrombosis, embolism, etc.
XX
XX Example 2; Page 20; 112pp; English.
PS
XX The polypeptide C was one of three polypeptides which were
CC synthesised based on amino acids 406 to 444 of human thrombomodulin.
CC (see R22001-R22002 for polypeptides A and B, respectively). The
CC polypeptides were analysed for their ability to inhibit binding of
CC thrombin to thrombomodulin. Polypeptides A and B had no binding
CC activity, while polypeptide C had an inhibition constant of
CC 95 microm. See also R22004, R22013-R22022 and Q22200-Q22203.
CC
XX Sequence 19 AA:
SQ
Query Match 39.6%; Score 42; DB 13; Length 19;
Best Local Similarity 56.2%; Pred. NO. 1;
Matches 9; Conservative 1; Mismatches 4; Indels 2; Gaps 1;
OY 1 CPTGTALDDG--VTDV 14
Db 2 cpegylddgtfictdi 17
RESULT 9
R41802
ID R41802 standard; peptide; 19 AA.
XX
AC R41802;
XX
XX 30-MAR-1994 (first entry)
DT
XX Thrombomodulin peptide C.
DE
XX Transformation; fungus; blood coagulation; prevention; platelet;
KW aggregation; thrombolytic activity; thrombolysis; human thrombomodulin;
KW site-directed mutagenesis; promotion; protein C; activation; thrombin.
XX
XX Synthetic.
OS
XX JP05213998-A.
PN
XX 24-AUG-1993.
PD
XX 02-AUG-1991; 91JP-0282369.
PF
XX 03-AUG-1990; 90JP-0204978.
PR 30-JUL-1991; 91JP-0189984.
XX
XX (ASAH ) ASAH CHEM IND CO LTD.
PA
XX WPI; 1993-299652/38.
DR
XX Novel polypeptide obtd. by culturing transformed fungus - having
PT blood coagulation preventing, platelet aggregation preventing and
PT thrombolytic activities
XX
XX Example 2; Page 13; 65pp; Japanese.
PS
XX Peptides A, B and C (R41800-R41802) are described in example 2.
CC

```

```

CC Novel polypeptides, obtd. by culturing transformed fungus, have
CC blood coagulation preventing, platelet aggregation preventing
CC and thrombolytic activities.
CC In an example, plasmid M13mp19TMD3 (constructed from pSV2TMJ2
CC (ATCC 67238) contg. a human thrombomodulin sequence) was subjected
CC to site directed mutagenesis to prepare pSV2TMJ7. Plasmid pSV2TMJ7
CC was transfected to COS-1 cells. The activity of promoting protein C
CC activation by thrombin of the peptide produced by the transformed
CC COS-1 cell was measured. The amt. of the peptide was determined.
CC
XX Sequence 19 AA:
SQ
Query Match 39.6%; Score 42; DB 14; Length 19;
Best Local Similarity 56.2%; Pred. NO. 1;
Matches 9; Conservative 1; Mismatches 4; Indels 2; Gaps 1;
OY 1 CPTGTALDDG--VTDV 14
Db 2 cpegylddgtfictdi 17
RESULT 10
R41798
ID R41798 standard; peptide; 20 AA.
XX
AC R41798;
XX
XX 30-MAR-1994 (first entry)
DT
XX Thrombomodulin polypeptide.
DE
XX Transformation; fungus; blood coagulation; prevention; platelet;
KW aggregation; thrombolytic activity; thrombolysis; human thrombomodulin;
KW site-directed mutagenesis; promotion; protein C; activation; thrombin.
XX
XX Synthetic.
OS
XX JP05213998-A.
PN
XX 24-AUG-1993.
PD
XX 02-AUG-1991; 91JP-0282369.
PF
XX 03-AUG-1990; 90JP-0204978.
PR 30-JUL-1991; 91JP-0189984.
XX
XX (ASAH ) ASAH CHEM IND CO LTD.
PA
XX WPI; 1993-299652/38.
DR
XX Novel polypeptide obtd. by culturing transformed fungus - having
PT blood coagulation preventing, platelet aggregation preventing and
PT thrombolytic activities
XX
XX Claim 1; Page 2; 65pp; Japanese.
PS
XX The polypeptide, obtd. by culturing transformed fungus, has blood
CC coagulation preventing, platelet aggregation preventing and
CC thrombolytic activities.
CC In an example, plasmid M13mp19TMD3 (constructed from pSV2TMJ2
CC (ATCC 67238) contg. a human thrombomodulin sequence) was subjected

```

CC to site directed mutagenesis to prepare pSV2TM7. Plasmid pSV2TM7
 CC was transfected to COS-1 cells. The activity of promoting protein C
 CC activation by thrombin of the peptide produced by the transformed
 CC COS-1 cell was measured. The amt. of the peptide was determined.
 CC
 SO Sequence 20 AA:

Query Match 39.6%; Score 42; DB 14; Length 20;
 Best Local Similarity 56.2%; Pred. No. 1.1;
 Matches 9; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

OY 1 CPTGTALDDG--VNDV 14
 1111111111
 Db 3 cpeyllddggfictdi 18

RESULT 11

W76685 ID W76685 standard; Protein; 70 AA.

XX AC W76685;

XX DF 05-JAN-1999 (first entry)

XX DE Elapidae modified dendroaspin protein fragment Den-Tm.

XX KW Dendroaspin; snake venom; clotting cascade; anticoagulant; platelet;
 KW integrin binding; injury; blood; cell migration; thrombosis; inhibitor;
 KW proliferation; signal transduction; regulator; coagulation; treatment;
 KW prophylactic; artery; vein; wall thickening; myocardial infarction;
 KW retinal neovascularisation; dysregulated apoptosis; tumorigenesis;
 KW leukocyte recruitment; immune system; tissue fibrosis.

OS Elapidae.
 OS Synthetic.

XX PN W09842834-A1.

XX PD 01-OCT-1998.

XX PF 20-MAR-1998; 98WO-GB00848.

XX PR 20-MAR-1997; 97GB-0005787.

XX PA (THRO-) THROMBOSIS RES INST.

XX PI Authi K, Kakkar V, Lu X, Scully MF;

XX DR WPI; 1998-542278/46.

XX PT New hybrid dendroaspin polypeptide(s) - used for treating, e.g.
 PT thrombosis, myocardial infarction, dysregulated apoptosis, abnormal
 PT cell migration and immune system activation

XX PS Claim 6; Fig 3C; 59pp; English.

XX W76645-W76688 represent modified dendroaspin protein fragments isolated
 CC from snake venom. When dendroaspin is modified to incorporate further
 CC functional amino acid sequence, e.g. active portions or motifs of
 CC agonists, antagonists or inhibitors of factors in the clotting cascade,
 CC the resulting molecules are particularly useful as anticoagulants. The
 CC molecules have an integrin binding activity which when administered in
 CC vivo results in the binding of the molecules to platelets thereby
 CC inhibiting the aggregation of the platelets at sites of injury. Non-wild
 CC type dendroaspin domains provide secondary, optionally further
 CC functionality, e.g. antithrombotic action, inhibiting cell migration and
 CC proliferation and regulating signal transduction. Such variants have bi-
 CC or multifunctional activities against blood coagulation, particularly
 CC thrombus formation and arterial/venous wall thickening at the sites of
 CC injury. The variants may have activities against leukocyte recruitment,
 CC immune system activation, tissue fibrosis and tumorigenesis. The
 CC polypeptides can be used for the treatment or prophylaxis of a disease

CC associated with thrombosis, e.g. myocardial infarction, retinal
 CC neovascularisation, endothelial injury, dysregulated apoptosis, abnormal
 CC cell migration, leukocyte recruitment, immune system activation, tissue
 CC fibrosis or tumorigenesis.
 CC
 SO Sequence 70 AA:

Query Match 39.6%; Score 42; DB 19; Length 70;
 Best Local Similarity 56.2%; Pred. No. 4.8;
 Matches 9; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

OY 1 CPTGTALDDG--VNDV 14
 1111111111
 Db 53 cpeyllddggfictdi 68

RESULT 12

R41799 ID R41799 standard; peptide; 114 AA.

XX AC R41799;

XX DF 30-MAR-1994 (first entry)

XX DE Thrombomodulin polypeptide.

XX KW Transformation; fungus; blood coagulation; prevention; platelet;
 KW aggregation; thrombolytic activity; thrombolytics; human thrombomodulin;
 KW site-directed mutagenesis; promotion; protein C; activation; thrombin.
 XX OS Synthetic.

XX FH Key Location/Qualifiers

FT FT MISC-difference 1 /note= "the N-terminal X may be an amino acid residue
 FT or a peptide residue consisting of only Asp or
 FT Glu and Glu (gamma-carboxyglutamic acid) or a
 FT peptide residue consisting of a combination
 FT of at least two of Asp, Glu or Glu"
 XX

XX PN JP05213998-A.

XX PD 24-AUG-1993.

XX PF 02-AUG-1991; 91JP-0282369.

XX PR 03-AUG-1990; 90JP-0204978.

XX PR 30-JUL-1991; 91JP-0189984.

XX PA (ASAH) ASAH CHEM IND CO LTD.

XX DR WPI; 1993-299652/38.

XX PT Novel polypeptide obtd. by culturing transformed fungus - having
 PT blood coagulation preventing, platelet aggregation preventing and
 PT thrombolytic activities

XX PS Claim 2; Page 2; 65pp; Japanese.

XX The polypeptide, obtd. by culturing transformed fungus, has blood
 CC coagulation preventing, platelet aggregation preventing and
 CC thrombolytic activities.

CC In an example, plasmid M13mp19TM3 (constructed from pSV2TM2
 CC (ATCC 67238) contg. a human thrombomodulin sequence) was subjected
 CC to site directed mutagenesis to prepare pSV2TM7. Plasmid pSV2TM7
 CC was transfected to COS-1 cells. The activity of promoting protein C
 CC activation by thrombin of the peptide produced by the transformed
 CC COS-1 cell was measured. The amt. of the peptide was determined.

SO Sequence 114 AA;

Query Match 39.6%; Score 42; DB 14; Length 114;
 Best Local Similarity 56.2%; Pred. NO. 8.6;
 Matches 9; Conservative 1; Mismatches 4; Indels 2; Gaps 1;
 OY 1 CPTGTALDDG--VTDV 14
 || | |||| |
 Db 61 cpegylddgtfictdi 76

RESULT 13

R84184.
 ID R84184 standard; protein; 114 AA.

AC R84184;

DT 05-JUN-1996 (first entry)

DE Human derived thrombomodulin fragment.

KW Human thrombomodulin; fragment; hepatopathy; fulminant hepatitis.

OS Homo sapiens.

PM W09528953-A1.

PD 02-NOV-1995.

PE 10-APR-1995; 95WO-JP00704.

PR 20-APR-1994; 94JP-0081196.

PA (ASAHI) ASAHI KASEI KOGYO KK.

PI Fujiwara K, Mochida S;

DR WPI: 1995-382844/49.

PT Hepatopathy treatment pref. comprising human derived thrombomodulin

PT - useful against, e.g, fulminant hepatitis

PS Claim 2; Page 17; 29pp; Japanese.

CC The human derived thrombomodulin R84185, and its fragment R84184
 CC can be used in the treatment of hepatopathy, including fulminant
 CC hepatitis.

SO Sequence 114 AA;

Query Match 39.6%; Score 42; DB 16; Length 114;
 Best Local Similarity 56.2%; Pred. NO. 8.6;

Matches 9; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

OY 1 CPTGTALDDG--VTDV 14
 || | |||| |

Db 61 cpegylddgtfictdi 76

RESULT 14

W01599
 ID W01599 standard; protein; 114 AA.

AC W01599;

DT 17-MAR-1997 (first entry)

DE Thrombomodulin TMD123 protein.

KW Thrombomodulin; nervous interference; treatment; medulla injury.

OS Homo sapiens.

PN JP08283174-A.

XX 29-OCT-1996.
 PD 11-APR-1995; 95JP-0085202.

PF 11-APR-1995; 95JP-0085202.

PR 11-APR-1995; 95JP-0085202.

PA (ASAHI) ASAHI KASEI KOGYO KK.

DR WPI: 1997-017314/02.

PT Treating agent for nervous interference after medulla injury -

PT contains thrombomodulin as active component

PS Claim 2; Page 4-5; 6pp; Japanese.

CC W01599-600 are fragments of human thrombomodulin. They can be used in

CC treatment of nervous interference after medulla injury. In an example, a

CC model of medulla injury was prepd. by compressing a male Wistar rat of

CC 200-250 g body wt., with a 20g weight for 20 mins.. TMD123 dissolved in

CC physiological saline water was injected into the tail vein 30 mins.

CC before injury. Movement function was evaluated 24 hrs. after injury. MPO

CC activity in medulla tissue was measured 3 hrs. after injury as an index

CC of neutrophilic accumulation. The Tarlov score was 1.50 compared to 0.77

CC for the control, and complete paralysis was 15.4 percent compared to

CC 35.3 percent for the control.

SO Sequence 114 AA;

Query Match 39.6%; Score 42; DB 18; Length 114;
 Best Local Similarity 56.2%; Pred. NO. 8.6;

Matches 9; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

OY 1 CPTGTALDDG--VTDV 14
 || | |||| |

Db 61 cpegylddgtfictdi 76

RESULT 15

R07074
 ID R07074 standard; protein; 115 AA.

AC R07074;

DT 18-FEB-1991 (first entry)

DE Anticoagulant and anti-platelet aggregation agent.

KW Gestosis; thrombus; catheter.

PM JP02255699-A.

PD 16-OCT-1990.

PR 28-MAR-1989; 89JP-0074009.

PA (ASAHI) ASAHI CHEMICAL IND KK.

DR WPI: 1990-352835/47.

PT N-PSDB; Q06574.

PT New 115-unit polypeptide and DNA encoding it - is anticoagulant

PT and anti-platelet aggregation agent

PS Claim 1; Page 827; 19pp; Japanese.

CC Gene product may be expressed from a vector transformed by the

CC sequence encoding it, and has anticoagulant and anti-platelet

CC aggregation activity as a result of combining with thrombin.

CC The product is effective in the treatment of diseases of the

CC circulatory organs and gestosis, and may be combined with artificial
CC blood vessels, internal organs and catheters.
XX
SQ sequence 115 AA;

Query Match 39.6%; Score 42; DB 11; Length 115;
Best Local Similarity 56.2%; Pred. No. 8.7;
Matches 9; Conservative 1; Mismatches 4; Indels 2; Gaps 1;
QY 1 CPTGTALDDG--VTDV 14
1111111111
Db 62 cpegyilddgfictdi 77

Search completed: March 6, 2001, 12:49:34
Job time: 95 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:53:49 ; Search time 57.76 Seconds
(without alignments)
6.218 Million cell updates/sec

Title: US-09-196-161d-5

Perfect score: 106

Sequence: 1 CPTGTALDDGVTVDFDRSAA 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43.4	362	2	US-08-846-762-14	Sequence 14, Appl
2	43.4	362	2	US-08-846-762-83	Sequence 83, Appl
3	44	41.5	355	2 US-08-641-314C-2	Sequence 2, Appl
4	44	41.5	424	1 US-08-419-414-2	Sequence 2, Appl
5	43	40.6	2183	3 US-08-746-111-5	Sequence 5, Appl
6	42	39.6	19	1 US-08-261-206A-2	Sequence 2, Appl
7	42	39.6	19	1 US-08-261-206A-27	Sequence 27, Appl
8	42	39.6	114	2 US-08-733-564-1	Sequence 1, Appl
9	42	39.6	115	1 US-08-312-870-9	Sequence 9, Appl
10	42	39.6	275	1 US-08-312-870-7	Sequence 7, Appl
11	42	39.6	446	1 US-08-307-444A-5	Sequence 5, Appl
12	42	39.6	456	1 US-08-587-389-5	Sequence 7, Appl
13	42	39.6	456	1 US-08-307-444A-3	Sequence 3, Appl
14	42	39.6	456	1 US-08-587-444A-4	Sequence 4, Appl
15	42	39.6	456	1 US-08-587-389-3	Sequence 4, Appl
16	42	39.6	456	1 US-08-587-389-4	Sequence 4, Appl
17	42	39.6	475	1 US-08-307-444A-1	Sequence 1, Appl
18	42	39.6	475	1 US-08-307-444A-2	Sequence 2, Appl
19	42	39.6	475	1 US-08-587-389-1	Sequence 1, Appl
20	42	39.6	475	1 US-08-587-389-2	Sequence 2, Appl
21	42	39.6	476	1 US-08-014-723-1	Sequence 1, Appl
22	42	39.6	476	1 US-08-014-723-2	Sequence 1, Appl
23	42	39.6	476	1 US-08-014-723-18	Sequence 18, Appl
24	42	39.6	476	1 US-08-110-011A-1	Sequence 1, Appl
25	42	39.6	476	1 US-08-110-011A-2	Sequence 2, Appl
26	42	39.6	476	1 US-08-110-011A-18	Sequence 18, Appl
27	42	39.6	494	1 US-08-014-723-14	Sequence 14, Appl
28	42	39.6	494	1 US-08-014-723-16	Sequence 16, Appl

29	42	39.6	494	1	US-08-110-011A-14	Sequence 14, Appl
30	42	39.6	494	1	US-08-110-011A-16	Sequence 16, Appl
31	42	39.6	497	1	US-08-312-870-3	Sequence 3, Appl
32	42	39.6	498	2	US-08-733-564-2	Sequence 2, Appl
33	42	39.6	572	5	5256770-7	Patent No. 5256770
34	42	39.6	575	1	US-08-261-206A-59	Sequence 59, Appl
35	42	39.6	575	1	US-08-312-870-1	Sequence 1, Appl
36	42	39.6	575	1	US-08-170-290A-54	Sequence 54, Appl
37	42	39.6	575	5	5466668-6	Patent No. 5466668
38	40	37.7	31	1	US-08-190-802A-242	Sequence 242, App
39	40	37.7	31	1	US-08-190-802A-260	Sequence 260, App
40	40	37.7	111	1	US-08-288-728-4	Sequence 4, Appl
41	40	37.7	798	1	US-08-190-802A-64	Sequence 64, Appl
42	40	37.7	798	1	US-08-190-802A-68	Sequence 68, Appl
43	40	37.7	798	2	US-08-308-818-2	Sequence 2, Appl
44	39	36.8	342	1	US-08-087-797-3	Sequence 3, Appl
45	39	36.8	397	5	5495001-7	Patent No. 5495001

ALIGNMENTS

```

RESULT 1
US-08-846-762-14
; Sequence 14, Application US/08846762A
; Patent No. 5994072
; GENERAL INFORMATION:
; APPLICANT: Lam, Joseph S.
; APPLICANT: Burrows, Lori
; APPLICANT: Charter, Deborah
; TITLE OF INVENTION: No. 5994072el Proteins Involved in the Synthesis and Assembly
; FILE REFERENCE: 6580-089
; CURRENT FILING DATE: 1997-04-30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 14
; LENGTH: 362
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-08-846-762-14

Query Match 43.4% Score 46: DB 2: Length 362:
Best Local Similarity 45.0% Pred. No. 4.1;
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 CPTGTALDDGVTVDFDRSAA 20
DB 146 CPTRTALDNLKNEGRKKA 165

RESULT 2
US-08-846-762-83
; Sequence 83, Application US/08846762A
; Patent No. 5994072
; GENERAL INFORMATION:
; APPLICANT: Lam, Joseph S.
; APPLICANT: Burrows, Lori
; APPLICANT: Charter, Deborah
; TITLE OF INVENTION: No. 5994072el Proteins Involved in the Synthesis and Assembly
; FILE REFERENCE: 6580-089
; CURRENT FILING DATE: 1997-04-30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 83
; LENGTH: 362
; TYPE: PRT

```

ORGANISM: *Pseudomonas aeruginosa*
US-08-846-762-83

Query Match	43.48;	Score 46;	DB 2;	Length 362;
Best Local Similarity	45.08;	Pred. NO. 4.1;		
Matches	9;	Conservative	4;	Mismatches 7;
				Indels

Qy 1 CPTGTALDDGVTDVFDPSAA 20
||| |:| : |:| ||
Db 146 CPTRAVDNLKNEGFERKAA 165

RESULT 3
US-08-641-314C-2
; Sequence 2, Application US/08641314C
; Patent No. 5977440
GENERAL INFORMATION:

1 GENERAL INFORMATION:
2 APPLICANT: LUTHE, DAWN S.
3 APPLICANT: WILLIAMS, W. P.
4 APPLICANT: BINGHUA, JIANG
5 APPLICANT: PECHAN, TIBOR
6 TITLE OF INVENTION: DNA MOLECULE ENCODING A 33 KD CYSTEINE
7 TITLE OF INVENTION: PROTEINASE AND ITS USE IN TRANSFORMING PLANTS TO PROVIDE
8 TITLE OF INVENTION: INSECT RESISTANCE
9 NUMBER OF SEQUENCES: 11
10 CORRESPONDENCE ADDRESS: *****
11 *****
12 *****
13 *****
14 *****
15 *****
16 *****
17 *****
18 *****
19 *****
20 *****
21 *****
22 *****
23 *****
24 *****
25 *****
26 *****
27 *****
28 *****
29 *****
30 *****
31 *****
32 *****
33 *****
34 *****
35 *****
36 *****
37 *****
38 *****
39 *****
40 *****
41 *****
42 *****
43 *****
44 *****
45 *****
46 *****
47 *****
48 *****
49 *****
50 *****
51 *****
52 *****
53 *****
54 *****
55 *****
56 *****
57 *****
58 *****
59 *****
60 *****
61 *****
62 *****
63 *****
64 *****
65 *****
66 *****
67 *****
68 *****
69 *****
70 *****
71 *****
72 *****
73 *****
74 *****
75 *****
76 *****
77 *****
78 *****
79 *****
80 *****
81 *****
82 *****
83 *****
84 *****
85 *****
86 *****
87 *****
88 *****
89 *****
90 *****
91 *****
92 *****
93 *****
94 *****
95 *****
96 *****
97 *****
98 *****
99 *****
100 *****
101 *****
102 *****
103 *****
104 *****
105 *****
106 *****
107 *****
108 *****
109 *****
110 *****
111 *****
112 *****
113 *****
114 *****
115 *****
116 *****
117 *****
118 *****
119 *****
120 *****
121 *****
122 *****
123 *****
124 *****
125 *****
126 *****
127 *****
128 *****
129 *****
130 *****
131 *****
132 *****
133 *****
134 *****
135 *****
136 *****
137 *****
138 *****
139 *****
140 *****
141 *****
142 *****
143 *****
144 *****
145 *****
146 *****
147 *****
148 *****
149 *****
150 *****
151 *****
152 *****
153 *****
154 *****
155 *****
156 *****
157 *****
158 *****
159 *****
160 *****
161 *****
162 *****
163 *****
164 *****
165 *****
166 *****
167 *****
168 *****
169 *****
170 *****
171 *****
172 *****
173 *****
174 *****
175 *****
176 *****
177 *****
178 *****
179 *****
180 *****
181 *****
182 *****
183 *****
184 *****
185 *****
186 *****
187 *****
188 *****
189 *****
190 *****
191 *****
192 *****
193 *****
194 *****
195 *****
196 *****
197 *****
198 *****
199 *****
200 *****
201 *****
202 *****
203 *****
204 *****
205 *****
206 *****
207 *****
208 *****
209 *****
210 *****
211 *****
212 *****
213 *****
214 *****
215 *****
216 *****
217 *****
218 *****
219 *****
220 *****
221 *****
222 *****
223 *****
224 *****
225 *****
226 *****
227 *****
228 *****
229 *****
230 *****
231 *****
232 *****
233 *****
234 *****
235 *****
236 *****
237 *****
238 *****
239 *****
240 *****
241 *****
242 *****
243 *****
244 *****
245 *****
246 *****
247 *****
248 *****
249 *****
250 *****
251 *****
252 *****
253 *****
254 *****
255 *****
256 *****
257 *****
258 *****
259 *****
260 *****
261 *****
262 *****
263 *****
264 *****
265 *****
266 *****
267 *****
268 *****
269 *****
270 *****
271 *****
272 *****
273 *****
274 *****
275 *****
276 *****
277 *****
278 *****
279 *****
280 *****
281 *****
282 *****
283 *****
284 *****
285 *****
286 *****
287 *****
288 *****
289 *****
290 *****
291 *****
292 *****
293 *****
294 *****
295 *****
296 *****
297 *****
298 *****
299 *****
300 *****
301 *****
302 *****
303 *****
304 *****
305 *****
306 *****
307 *****
308 *****
309 *****
310 *****
311 *****
312 *****
313 *****
314 *****
315 *****
316 *****
317 *****
318 *****
319 *****
320 *****
321 *****
322 *****
323 *****
324 *****
325 *****
326 *****
327 *****
328 *****
329 *****
330 *****
331 *****
332 *****
333 *****
334 *****
335 *****
336 *****
337 *****
338 *****
339 *****
340 *****
341 *****
342 *****
343 *****
344 *****
345 *****
346 *****
347 *****
348 *****
349 *****
350 *****
351 *****
352 *****
353 *****
354 *****
355 *****
356 *****
357 *****
358 *****
359 *****
360 *****
361 *****
362 *****
363 *****
364 *****
365 *****
366 *****
367 *****
368 *****
369 *****
370 *****
371 *****
372 *****
373 *****
374 *****
375 *****
376 *****
377 *****
378 *****
379 *****
380 *****
381 *****
382 *****
383 *****
384 *****
385 *****
386 *****
387 *****
388 *****
389 *****
390 *****
391 *****
392 *****
393 *****
394 *****
395 *****
396 *****
397 *****
398 *****
399 *****
400 *****
401 *****
402 *****
403 *****
404 *****
405 *****
406 *****
407 *****
408 *****
409 *****
410 *****
411 *****
412 *****
413 *****
414 *****
415 *****
416 *****
417 *****
418 *****
419 *****
420 *****
421 *****
422 *****
423 *****
424 *****
425 *****
426 *****
427 *****
428 *****
429 *****
430 *****
431 *****
432 *****
433 *****
434 *****
435 *****
436 *****
437 *****
438 *****
439 *****
440 *****
441 *****
442 *****
443 *****
444 *****
445 *****
446 *****
447 *****
448 *****
449 *****
450 *****
451 *****
452 *****
453 *****
454 *****
455 *****
456 *****
457 *****
458 *****
459 *****
460 *****
461 *****
462 *****
463 *****
464 *****
465 *****
466 *****
467 *****
468 *****
469 *****
470 *****
471 *****
472 *****
473 *****
474 *****
475 *****
476 *****
477 *****
478 *****
479 *****
480 *****
481 *****
482 *****
483 *****
484 *****
485 *****
486 *****
487 *****
488 *****
489 *****
490 *****
491 *****
492 *****
493 *****
494 *****
495 *****
496 *****
497 *****
498 *****
499 *****
500 *****
501 *****
502 *****
503 *****
504 *****
505 *****
506 *****
507 *****
508 *****
509 *****
510 *****
511 *****
512 *****
513 *****
514 *****
515 *****
516 *****
517 *****
518 *****
519 *****
520 *****
521 *****
522 *****
523 *****
524 *****
525 *****
526 *****
527 *****
528 *****
529 *****
530 *****
531 *****
532 *****
533 *****
534 *****
535 *****
536 *****
537 *****
538 *****
539 *****
540 *****
541 *****
542 *****
543 *****
544 *****
545 *****
546 *****
547 *****
548 *****
549 *****
550 *****
551 *****
552 *****
553 *****
554 *****
555 *****
556 *****
557 *****
558 *****
559 *****
560 *****
561 *****
562 *****
563 *****
564 *****
565 *****
566 *****
567 *****
568 *****
569 *****
570 *****
571 *****
572 *****
573 *****
574 *****
575 *****
576 *****
577 *****
578 *****
579 *****
580 *****
581 *****
582 *****
583 *****
584 *****
585 *****
586 *****

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/641,314C
FILING DATE: 30-Apr-1996

```

1  ATTORNEY/AGENT INFORMATION:
2  NAME:  KELBER, STEVEN B.
3  REGISTRATION NUMBER:  30,073
4  REFERENCE/DOCKET NUMBER:  2343-045-27
5  TELECOMMUNICATION INFORMATION:
6  TELEPHONE:  703-413-3000
7  TELEFAX:  703-413-3220
8  INFORMATION FOR SEQ ID NO:  2:
9  SEQUENCE CHARACTERISTICS:
10     LENGTH:  355 amino acids
11     TYPE:  amino acid
12     STRANDEDNESS:  single
13     TOPOLOGY:  linear
14     MOLECULE TYPE:  protein
15     OS-08-641-314C-2

```

Query Match	41.5%;	Score 44;	DB 2;	Length 355;
Best Local Similarity	69.2%;	Pred. No. 8.7;		
Matches	9;	Conservative	1;	Mismatches 3;
				Indels 0;
				Gaps 0;

Qy	2	PTGTALDDGVTDV	14
Db	267	PCGTSLDHGVTAV	279

RESULT 4
US-08-419-414-2
; Sequence 2, Application US/08419414
; Patent No. 5753787

GENERAL INFORMATION:

APPLICANT: Hawdon, John M.
 APPLICANT: Hotez, Peter J.
 APPLICANT: Jones, Brian F.
 TITLE OF INVENTION: Hookworm Vaccine
 NUMBER OF SEQUENCES: 16
 CORRESPONDENCE ADDRESS:

```

ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/419,414

```

CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patricia L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: YU1133
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8795
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 424 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
HYDROTHERICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:

ORGANISM: Ancylostoma caninum
DEVELOPMENTAL STAGE: Larva
-08-419-414-2

Query Match	41.58;	Score 44;	DB 1;	Length 424;
Best Local Similarity	46.7%;	Pred. No. 11;		
Matches	7;	Conservative	2;	Mismatches 6;
				Indels 0;
				Gaps 0;

```

1 CPTGALDDGVTDVF 15
  ||:|:|||||
227 CPSNTGMTDSVRDTF 241

```

SUPLY 5
-08-746-111-5
Sequence 5, Application US/08746111
Patent No. 6066778
GENERAL INFORMATION:

APPLICANT: Ginsburg, David
 APPLICANT: Cui, Jisong
 TITLE OF INVENTION: Compositions And Methods For Screening
 TITLE OF INVENTION: Compounds For Anticoagulant Activity
 NUMBER OF SEQUENCES: 54

ADDRESSEE: Medgen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/746,111
FILING DATE: 06-NOV-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: UM-02536
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2183 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-746-111-5

Query Match 40.6% Score 43; DB 3; Length 2183;
Best Local Similarity 41.2%; Pred. No. 1.1e+02;
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CPTGTALDDGVTDFDR 17
1 11 11 11
Db 192 CKGTTFEDGTQKMFDR 208

RESULT 6
US-08-261-206A-2
Sequence 2, Application US/08261206A
Patent No. 5574007
GENERAL INFORMATION:
APPLICANT: Zushi, Mitichitaka
APPLICANT: Gomi, Komakazu
APPLICANT: Yamamoto, Shuji
APPLICANT: Suzuki, Koji
APPLICANT: Matsuda, Akio
TITLE OF INVENTION: A polypeptide Capable of Interacting
NUMBER OF SEQUENCES: 80
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 301 N. Washington St.
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22046-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/261,206A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/740,492
FILING DATE: 03-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Svensson, Leonard R.
REGISTRATION NUMBER: 30330
REFERENCE/DOCKET NUMBER: 216-275P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-241-1300
TELEFAX: 703-241-2848
TELEX: 248345
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 19 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..19
OTHER INFORMATION: /label= peptide
OTHER INFORMATION: /note= "Peptide I, fragment of thrombin binding
OTHER INFORMATION: polypeptide."
US-08-261-206A-2

Query Match 39.6% Score 42; DB 1; Length 19;
Best Local Similarity 56.2%; Pred. No. 0.61;
Matches 9; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

Qy 1 CPTGTALDDG-VTDV 14
1 1 1 1 1 1 1
Db 2 CPEGTLDDGFICTDI 17

RESULT 7
US-08-261-206A-27
Sequence 27, Application US/08261206A
Patent No. 5574007
GENERAL INFORMATION:
APPLICANT: Zushi, Mitichitaka
APPLICANT: Gomi, Komakazu
APPLICANT: Yamamoto, Shuji
APPLICANT: Suzuki, Koji
APPLICANT: Matsuda, Akio
TITLE OF INVENTION: A polypeptide Capable of Interacting
NUMBER OF SEQUENCES: 80
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 301 N. Washington St.
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22046-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/261,206A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/740,492
FILING DATE: 03-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Svensson, Leonard R.
REGISTRATION NUMBER: 30330
REFERENCE/DOCKET NUMBER: 216-275P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-241-1300
TELEFAX: 703-241-2848
TELEX: 248345
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 19 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
FEATURE:
NAME/KEY: Peptide

LOCATION: 1..19
OTHER INFORMATION: /label= peptide
OTHER INFORMATION: /note= "synthetic peptide from thrombomodulin for
OTHER INFORMATION: study of thrombin binding site."
US-08-261-206A-27

Query Match 39.6%; Score 42; DB 1; Length 19;
Best Local Similarity 56.2%; Pred. No. 0.61;
Matches 9; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

QY 1 CPTGTALDDG--VTDV 14
Db 2 CPEGYILDDGFCITDI 17

RESULT 8
US-08-733-564-1

Sequence 1, Application US/08733564
Patent No. 5916874
GENERAL INFORMATION:
APPLICANT: FUJIWARA, Kenji
APPLICANT: MOCHIDA, Satoshi
TITLE OF INVENTION: METHOD FOR TREATING LIVER INJURY
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/733,564
FILING DATE: 18 OCTOBER 1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: SVENSSON, Leonard R.
REGISTRATION NUMBER: 30,330
REFERENCE/DOCKET NUMBER: 0216-0362P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 114 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-733-564-1

Query Match 39.6%; Score 42; DB 2; Length 114;
Best Local Similarity 56.2%; Pred. No. 5;
Matches 9; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

QY 1 CPTGTALDDG--VTDV 14
Db 61 CPEGYILDDGFCITDI 76

RESULT 9
US-08-312-870-9
Sequence 9, Application US/08312870
Patent No. 5639625
GENERAL INFORMATION:
APPLICANT: Carson, Craig W.

APPLICANT: Esmon, Charles T.
TITLE OF INVENTION: Method for Detecting Antibodies to
TITLE OF INVENTION: Thrombomodulin in Patients
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richards, Medlock & Andrews
STREET: 1201 Elm Street, Suite 4500
CITY: Dallas
STATE: Texas
COUNTRY: US
ZIP: 75270-2197

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/312,870
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hansen, Eugenia S.
REGISTRATION NUMBER: 31,966
REFERENCE/DOCKET NUMBER: OWRP B35150
TELECOMMUNICATION INFORMATION:
TELEPHONE: 214-939-4600
TELEFAX: 214-939-4500
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 115 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-312-870-9

Query Match 39.6%; Score 42; DB 1; Length 115;
Best Local Similarity 56.2%; Pred. No. 5;
Matches 9; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

QY 1 CPTGTALDDG--VTDV 14
Db 62 CPEGYILDDGFCITDI 77

RESULT 10
US-08-312-870-7
Sequence 7, Application US/08312870
Patent No. 5639625
GENERAL INFORMATION:
APPLICANT: Carson, Craig W.
APPLICANT: Esmon, Charles T.
TITLE OF INVENTION: Method for Detecting Antibodies to
TITLE OF INVENTION: Thrombomodulin in Patients
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richards, Medlock & Andrews
STREET: 1201 Elm Street, Suite 4500
CITY: Dallas
STATE: Texas
COUNTRY: US
ZIP: 75270-2197
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/312,870
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hansen, Eugenia S.

REGISTRATION NUMBER: 31,966
REFERENCE/DOCKET NUMBER: OMRF B35150
TELECOMMUNICATION INFORMATION:
TELEPHONE: 214-939-4500
TELEFAX: 214-939-4600
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 275 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-312-870-7

Query Match 39.6%; Score 42; DB 1; Length 275;
Best Local Similarity 56.2%; Pred. No. 14;
Matches 9; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

QY 1 CPTGALDDG--VTDV 14
||| ||||| ||:
Db 187 CPEGYLDDGFICTDI 202

RESULT 11
US-08-307-444A-5
Sequence 5, Application US/08307444A
Patent No. 5516659
GENERAL INFORMATION:
APPLICANT: NII, ATSUSHI
APPLICANT: MORISHITA, HIDEAKI
APPLICANT: UEMURA, AKIO
APPLICANT: MOCHIDA, EI
TITLE OF INVENTION: ANTICOAGULANT POLYPEPTIDES
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIEF & BERRIDGE
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/307,444A
FILING DATE: 19-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/835,436
FILING DATE: 26-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: OLIEF, JAMES A.
REGISTRATION NUMBER: 27,075
REFERENCE/DOCKET NUMBER: JAO 27706
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6400
TELEFAX: (703) 836-2787
TELEX: 90-1799 PTO ALEX
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 446 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-307-444A-5

Query Match 39.6%; Score 42; DB 1; Length 446;
Best Local Similarity 56.2%; Pred. No. 24;
Matches 9; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

QY 1 CPTGALDDG--VTDV 14
||| ||||| ||:
Db 409 CPEGYLDDGFICTDI 424

RESULT 12
US-08-587-389-5
Sequence 5, Application US/08587389
Patent No. 5695964
GENERAL INFORMATION:
APPLICANT: NII, ATSUSHI
APPLICANT: MORISHITA, HIDEAKI
APPLICANT: UEMURA, AKIO
APPLICANT: MOCHIDA, EI
TITLE OF INVENTION: TRUNCATED THROMBOMODULIN, RECOMBINANT
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIEF & BERRIDGE
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/587,389
FILING DATE: 17-JAN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/307,444
FILING DATE: 19-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: OLIEF, JAMES A.
REGISTRATION NUMBER: 27,075
REFERENCE/DOCKET NUMBER: JAO 27706
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6400
TELEFAX: (703) 836-2787
TELEX: 90-1799 PTO ALEX
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 446 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-587-389-5

Query Match 39.6%; Score 42; DB 1; Length 446;
Best Local Similarity 56.2%; Pred. No. 24;
Matches 9; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

QY 1 CPTGALDDG--VTDV 14
||| ||||| ||:
Db 409 CPEGYLDDGFICTDI 424

RESULT 13
US-08-307-444A-3
Sequence 3, Application US/08307444A
Patent No. 5516659

GENERAL INFORMATION:
APPLICANT: NII, ATSUSHI
APPLICANT: MORISHITA, HIDEAKI
APPLICANT: UEMURA, AKIO
APPLICANT: MOCHIDA, EI
TITLE OF INVENTION: ANTICOAGULANT POLYPEPTIDES
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/307,444A
FILING DATE: 19-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/835,436
FILING DATE: 26-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: OLIFF, JAMES A.
REGISTRATION NUMBER: 27,075
REFERENCE/DOCKET NUMBER: JAO 27706
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6400
TELEFAX: (703) 836-2787
TELEX: 90-1799 PTO ALEX
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-307-444A-3

Query Match 39.6%; Score 42; DB 1; Length 456;
Best Local Similarity 56.2%; Pred. No. 25;
Matches 9; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

QY 1 CPTGTALDDG--VTDV 14
||| ||| |||
Db 409 CPEGYILDDGFTCTDI 424

RESULT 14
US-08-307-444A-4
Sequence 4, Application US/08307444A
Patent No. 5516659
GENERAL INFORMATION:
APPLICANT: NII, ATSUSHI
APPLICANT: MORISHITA, HIDEAKI
APPLICANT: UEMURA, AKIO
APPLICANT: MOCHIDA, EI
TITLE OF INVENTION: ANTICOAGULANT POLYPEPTIDES
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/307,444A
FILING DATE: 19-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/835,436
FILING DATE: 26-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: OLIFF, JAMES A.
REGISTRATION NUMBER: 27,075
REFERENCE/DOCKET NUMBER: JAO 27706
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6400
TELEFAX: (703) 836-2787
TELEX: 90-1799 PTO ALEX
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-307-444A-4

Query Match 39.6%; Score 42; DB 1; Length 456;
Best Local Similarity 56.2%; Pred. No. 25;
Matches 9; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

QY 1 CPTGTALDDG--VTDV 14
||| ||| |||
Db 409 CPEGYILDDGFTCTDI 424

RESULT 15
US-08-587-389-3
Sequence 3, Application US/08587389
Patent No. 5695964
GENERAL INFORMATION:
APPLICANT: NII, ATSUSHI
APPLICANT: MORISHITA, HIDEAKI
APPLICANT: UEMURA, AKIO
APPLICANT: MOCHIDA, EI
TITLE OF INVENTION: TRUNCATED THROMBOMODULIN, RECOMBINANT
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/587,389
FILING DATE: 17-JAN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/307,444
FILING DATE: 19-SEP-1994
ATTORNEY/AGENT INFORMATION:

NAME: OLIFF, JAMES A.
 REGISTRATION NUMBER: 27,075
 REFERENCE/DOCKET NUMBER: JAO 27706
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-6400
 TELEFAX: (703) 836-2787
 TELE: 90-1799 PTO ALEX
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 456 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 US-08-587-389-3

Query Match 39.6%; Score 42; DB 1; Length 456;
 Best Local Similarity 56.2%; Pred. No. 25;
 Matches 9; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

OY 1 CPTGTAALDDG--VTDV 14
 ||| ||||| ||:
 Db 409 CPEGYILDDGFICTDI 424

Search completed: March 6, 2001, 12:53:49
 Job time: 349 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OK protein - protein search, using sw model

Run on: March 6, 2001, 12:50:50 ; Search time 70.34 Seconds
(without alignments)
19.306 Million cell updates/sec

Title: us-09-196-161d-5
Perfect score: 106
Sequence: 1 CPTGTALDDGVTVDFDRSAA 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues
Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_66: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106	100.0	395	2 A46031	Immobiliation sur
2	48	45.3	923	2 H64081	Arp-dependent heli
3	46	43.4	354	2 C83253	probable UDP-N-ace
4	46	43.4	466	2 H71207	hypothetical prote
5	45	42.5	387	2 E71711	probable O-sialogl
6	44	41.5	111	2 T49561	hypothetical prote
7	44	41.5	699	2 T15920	hypothetical prote
8	43	40.6	2180	2 T29764	hypothetical prote
9	43	40.6	2183	2 T42764	coagulation factor
10	42.5	40.1	255	2 T22863	hypothetical prote
11	42.5	40.1	372	2 T20650	hypothetical prote
12	42.5	40.1	527	2 T22867	hypothetical prote
13	42	39.6	87	2 T51449	hypothetical prote
14	42	39.6	254	2 S19390	probable membrane
15	42	39.6	305	1 S52775	hypothetical prote
16	42	39.6	374	2 T35581	probable O-sialogl
17	42	39.6	461	2 T05773	protein kinase hom
18	42	39.6	575	1 THH0B	thrombomodulin pre
19	42	39.6	1373	2 JEO095	gastric mucin MUC5
20	41	38.7	123	2 A83041	conserved hypotet
21	41	38.7	21	2 H69079	conserved hypotet
22	41	38.7	350	2 T37511	probable phosphor
23	41	38.7	378	1 B38178	tela protein - pla
24	41	38.7	399	2 G75332	acetyl-CoA acetyl
25	41	38.7	418	2 T40913	probable integral
26	41	38.7	487	2 T32341	hypothetical prote
27	41	38.7	529	1 DRECPH	putH bifunctional
28	41	38.7	741	2 T22466	hypothetical prote
29	41	38.7	765	1 T44946	transducer protein

30	41	38.7	1115	2 D69617	DNA polymerase III
31	41	38.7	1369	2 T32338	hypothetical prote
32	41	38.7	1451	2 T30821	P-glycoprotein - S
33	40.5	38.2	242	2 T50297	probable aminopept
34	40.5	38.2	395	2 D81146	aminotransferase,
35	40.5	38.2	395	2 E81877	probable aminotran
36	40.5	38.2	713	2 A35502	major surface-labe
37	40	37.7	63	2 A27536	Arpase inhibitor,
38	40	37.7	205	2 T14651	hypothetical prote
39	40	37.7	210	2 T14967	hypothetical prote
40	40	37.7	223	1 C57987	nitric protein - Bsc
41	40	37.7	298	2 B71872	hypothetical prote
42	40	37.7	311	2 S36619	protein kinase (EC
43	40	37.7	321	2 F75287	probable DNA polym
44	40	37.7	355	2 JC1249	peroxidase (EC 1.1
45	40	37.7	364	2 S34355	peroxidase (EC 1.1

ALIGNMENTS

```

RESULT 1
A46031
Immobiliation surface T-antigen precursor - Ichthyophthirius multifiliis (fragment)
C:Species: Ichthyophthirius multifiliis
C:Date: 29-Sep-1999 #sequence_revision=29-Sep-1999 #text_change 07-Dec-1999
C:Accession: A46031
R:Clark, T.G.; McGraw, R.A.; Dickerson, H.W.
Proc. Natl. Acad. Sci. U.S.A. 89, 6363-6367, 1992
A:Title: Developmental expression of surface antigen genes in the parasitic ciliate I
A:Reference number: A46031; MID:92435298
A:Accession: A46031
A:Molecule type: mRNA; protein
A:Residues: 1-395 <CLAS>
A:Cross-references: GB:M92907; NID:g3628568; PIDN:AAC36158.1; PID:g3628569
A:Note: The authors translated the codon UUG for residue 330 as Ile
A:Note: sequence extracted from NCBI backbone (NCBIN:108734, NCBI:P:108735); the seque
C:Genetics:
A:Genetic code: SGC5
C:Keywords: glycoprotein; surface antigen
P:2-395/Product: Immobiliation surface T-antigen #status experimental <MAT>.
F:156,191,245,281/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0% Score 106; DB 2; Length 395;
Best Local Similarity 100.0%; Pred. No. 6.5e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CPTGTALDDGVTVDFDRSAA 20
Db 87 CPTGTALDDGVTVDFDRSAA 106

RESULT 2
H64081
Arp-dependent helicase hepa homolog H10616 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 20-Sep-1999
C:Accession: H64081
R:Feilschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
J.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrman, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: H64081
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-923 <TIGR>
A:Cross-references: GB:U32744; GB:L42023; NID:g1573608; PIDN:AAC32275.1; PID:g1573610
C:Superfamily: unassigned DEAD/H box helicases; DEAD/H box helicase homology
C:Keywords: Arp; P-loop

```

F:175-552/Domain: DEAD/H box helicase homology <DEAD>
F:175-182/Region: nucleotide-binding motif A (P-loop)
F:274-279/Region: nucleotide-binding motif B
F:278-281/Region: DEAH motif

Query Match 45.3%; Score 48; DB 2; Length 923;
Best Local Similarity 44.4%; Pred. No. 17;
Matches 12; Conservative 2; Mismatches 5; Indels 8; Gaps 1;

OY 2 PTGTAL-----DDGVTDFDRSAA 20
||||| :||| |||||
Db 697 PTGTMVDPFPGKKEGVTVPDRRLA 723

RESULT 3
C83253
Probable UDP-N-acetylglucosamine 2-epimerase Wbpl PA3148 [imported] - Pseudomonas aeruginosa

C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Sep-2000
C:Accession: C83253

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laddig, K.; Lim, N.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: A82950
A:Accession: C83253

A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-354 <STO>

A:Cross-references: GB:AE004738; GB:AE004091; NID:g9949252; PIDN:MAG06536.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:

A:Gene: wbpI; PA3148

Query Match 43.4%; Score 46; DB 2; Length 354;
Best Local Similarity 45.0%; Pred. No. 13;
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 1 CPTGTALDDGVTDFDRSAA 20
||||| :||| |||||
Db 146 CPTNAIDNKNKEGFERKAA 165

RESULT 4

H71207

hypothetical protein PH1928 - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000

C:Accession: H71207

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekita, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushiida, N.; Oguchi, D.N. Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaeon

A:Reference number: A71000; MUID:98344137

A:Accession: H71207

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-466 <RAW>

A:Cross-references: GB:AP000007; NID:g3236134; PIDN:BAA31055.1; PID:g3258372

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by Genbank

C:Genetics:

A:Gene: PH1928

Query Match 43.4%; Score 46; DB 2; Length 466;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 6 ALDDGVTDFDR 17

Db 399 ALDEGIVDMDR 410
||||| :||| |||||

RESULT 5

E71711

Probable O-sialoglycoprotein endopeptidase (gcp) RP037 - Rickettsia prowazekii

C:Species: Rickettsia prowazekii

C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 20-Jun-2000

C:Accession: E71711

R:Anderson, S.G.E.; Zomorodipour, A.; Anderson, J.O.; Slichter-Ponten, T.; Alsmark, Nature 396, 133-140, 1998

A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.

A:Reference number: A71630; MUID:99039499

A:Accession: E71711

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-387 <AND>

A:Cross-references: GB:AJ235270; GB:AJ235269; NID:g3860572; PIDN:CAA14508.1; PID:g386

A:Experimental source: strain Madrid E

C:Genetics:

A:Gene: gcp; RP037

C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 42.5%; Score 45; DB 2; Length 387;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 4 GTALDDGVTDFDRSAA 19
||||| :||| |||||
Db 155 GTTIDDAVGETFDKVA 170

RESULT 6

T49561

hypothetical protein B208.90 [imported] - Neurospora crassa

C:Species: Neurospora crassa

C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 18-Aug-2000

C:Accession: T49561

R:Schulte, U.; Aign, V.; Hohnsbein, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu, submitted to the Protein Sequence Database, May 2000

A:Reference number: Z25022

A:Accession: T49561

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-111 <SCH>

A:Cross-references: EMBL:AL355930; GSPDB:GN00116; NCSP:B208.90

A:Experimental source: BAC clone B208; strain OR74A

C:Genetics:

A:Gene: NCSP:B208.90

A:Map position: 6

C:Superfamily: Neurospora crassa hypothetical protein B208.90

Query Match 41.5%; Score 44; DB 2; Length 111;
Best Local Similarity 61.5%; Pred. No. 8.4;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 CPTGTALDDGVTVD 13
||||| :||| |||||
Db 28 CPSAALDDGVTVD 40

RESULT 7

T15920

hypothetical protein EEBD8.10 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T15920

R:Chisoe, S.

Submitted to the EMBL Data Library, July 1995

A:Description: The sequence of C. elegans cosmid EEBD8.

A:Reference number: 218428
 A:Accession: T15920
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-699 <CHI>
 A:Cross-references: EMBL:U23484; NID:g733597; PID:g733607; PIDN:AAC46770.1; CESP:EEED8.1
 C:Genetics:
 A:Experimental source: strain Bristol N2
 A:Gene: CESP:EEED8.10
 A:Introns: 21/3; 63/2; 97/1; 137/3; 171/2; 198/3; 267/3; 463/2; 591/3; 662/1

Query Match 41.5%; Score 44; DB 2; Length 699;
 Best Local Similarity 58.8%; Pred. No. 53;
 Matches 10; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 4 GTALDDGVTVDFDRSA 20
 Db 87 GDAFDGGETDPDFDASIA 103

RESULT 8
 T29764
 hypothetical protein T21E3.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Aug-2000
 C:Accession: T29764
 R:Du, Z.; Le, T.T.
 submitted to the EMBL Data Library May 1997
 A:Description: The sequence of C. elegans cosmid T21E3.
 A:Reference number: Z20681
 A:Accession: T29764
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2180 <DUZ>
 A:Cross-references: EMBL:AF003133; PIDN:AB54138.1; GSPDB:GN00019; CESP:T21E3.3
 A:Experimental source: strain Bristol N2; clone T21E3
 C:Genetics:
 A:Gene: CESP:T21E3.3
 A:Map position: 1
 A:Introns: 400/2; 1608/3; 1644/1; 1674/3; 1686/2; 1731/2; 2073/1; 2098/2; 2146/3

C:Superfamily: LDL receptor ligand-binding repeat homology
 F:15-49/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F:51-85/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F:90-133/Domain: LDL receptor ligand-binding repeat homology <LDL3>
 F:149-185/Domain: LDL receptor ligand-binding repeat homology <LDL4>
 F:190-225/Domain: LDL receptor ligand-binding repeat homology <LDL5>
 F:228-272/Domain: LDL receptor ligand-binding repeat homology <LDL6>
 F:285-316/Domain: LDL receptor ligand-binding repeat homology <LDL7>
 F:323-359/Domain: LDL receptor ligand-binding repeat homology <LDL8>
 F:369-405/Domain: LDL receptor ligand-binding repeat homology <LDL9>
 F:816-866/Domain: LDL receptor ligand-binding repeat homology <LDL10>
 F:861-903/Domain: LDL receptor ligand-binding repeat homology <LDL11>
 F:908-943/Domain: LDL receptor ligand-binding repeat homology <LDL12>
 F:948-987/Domain: LDL receptor ligand-binding repeat homology <LDL13>
 F:993-1027/Domain: LDL receptor ligand-binding repeat homology <LDL14>
 F:1031-1063/Domain: LDL receptor ligand-binding repeat homology <LDL15>
 F:1126-1165/Domain: LDL receptor ligand-binding repeat homology <LDL16>
 F:1172-1209/Domain: LDL receptor ligand-binding repeat homology <LDL17>
 F:1213-1246/Domain: LDL receptor ligand-binding repeat homology <LDL18>

Query Match 40.6%; Score 43; DB 2; Length 2180;
 Best Local Similarity 64.3%; Pred. No. 2.4e+02;
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Oy 5 TALDDGVTVDFDRS 18
 Db 1189 TALCDGIRDCFDGS 1202

RESULT 9
 T42764

coagulation factor V - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000
 C:Accession: T42764
 R:Yang, T.L.; Cui, J.; Rehmertulle, A.; Yang, A.; Moussalli, M.; Kaufman, R.J.; Ginsbu
 Blood 91, 4593-4599, 1998
 A:Title: The structure and function of murine factor V and its inactivation by protei
 A:Reference number: 222270; MUID:98282202
 A:Accession: T42764
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2183 <YAN>
 A:Cross-references: EMBL:U52925; NID:g3219690; PID:g3219691; PIDN:AAC99553.1
 C:Function:
 A:Pathway: blood coagulation
 C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase
 C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasm
 F:350-682/Domain: ferroxidase repeat homology <FOX1>
 F:1541-1864/Domain: ferroxidase repeat homology <FOX2>

Query Match 40.6%; Score 43; DB 2; Length 2183;
 Best Local Similarity 41.2%; Pred. No. 2.4e+02;
 Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Oy 1 CPTGALDDGVTVDFDR 17
 Db 192 CKKGITLEDCTQKMPDK 208

RESULT 10
 T22863
 hypothetical protein F57G4.6 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T22863
 R:Ainscough, R.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z19629
 A:Accession: T22863
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-255 <WIL>
 A:Cross-references: EMBL:Z81554; PIDN:CAB04505.1; GSPDB:GN00023; CESP:F57G4.6
 A:Experimental source: clone F57G4
 C:Genetics:
 A:Gene: CESP:F57G4.6
 A:Map position: 5
 A:Introns: 31/3; 137/2; 174/1

Query Match 40.1%; Score 42.5; DB 2; Length 255;
 Best Local Similarity 50.0%; Pred. No. 33;
 Matches 8; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

Oy 1 CPTGALDDGVTVDFD 16
 Db 104 CPGSGGLDD-INSFD 118

RESULT 11
 T20650
 hypothetical protein F09G6.5 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
 C:Accession: T20650
 R:Mortimore, B.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z19305
 A:Accession: T20650
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-372 <WIL>

A:Cross-references: EMBL:Z81496; PIDN:CA804072.1; GSPDB:GN00023; CESP:F09C6.5
A:Experimental source: clone F09C6
C:Genetics:
A:Gene: CESP:F09C6.5
A:Map position: 5
A:Introns: 65/1; 84/2; 142/1; 246/3
C:Superfamily: antithrombin III

Query Match 40.1%; Score 42.5; DB 2; Length 372;
Best Local Similarity 55.6%; Pred. No. 48;
Matches 10; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

OY 4 GTALDD-GVTVDFRSAA 20
| | | | | | | | | |
Db 261 GAAIDDLGKDAFDMCAA 278

RESULT 12

T22867
hypothetical protein F5764.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T22867

R:Almscough, R.
submitted to the EMBL Data Library, November 1996

A:Reference number: Z19629

A:Accession: T22867

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-527 <WILL>

A:Cross-references: EMBL:Z81554; PIDN:CA804509.1; GSPDB:GN00023; CESP:F5764.7

A:Experimental source: clone F5764

C:Genetics:

A:Gene: CESP:F5764.7

A:Map position: 5

A:Introns: 6/3; 103/3; 209/2; 246/1; 400/3; 452/1

Query Match 40.1%; Score 42.5; DB 2; Length 527;
Best Local Similarity 50.0%; Pred. No. 68;
Matches 8; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

OY 1 CPTGTALDDGVTVDFD 16
| | | | | | | | | |
Db 176 CPSSGGLDD-IENSPD 190

RESULT 13

T51449
hypothetical protein F2G14.160 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000

C:Accession: T51449

R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mey
submitted to the Protein Sequence Database, August 2000

A:Reference number: Z25394

A:Accession: T51449

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-87 <SAT>

A:Cross-references: EMBL:AL391146

A:Experimental source: cultivar Columbia; BAC clone F2G14

C:Genetics:

A:Map position: 5

A:Introns: 36/2

A:Note: F2G14_160

Query Match 39.6%; Score 42; DB 2; Length 87;
Best Local Similarity 55.0%; Pred. No. 13;
Matches 11; Conservative 1; Mismatches 6; Indels 2; Gaps 1;

OY 2 PTG--TALDDGVTVDFRSA 19
| | | | | | | | | |
Db 65 PTGKRIALDDVDEDSFSSSTA 84

RESULT 14

S19390

probable membrane protein YCL005w - yeast (Saccharomyces cerevisiae)

C:Species: Saccharomyces cerevisiae

C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000

C:Accession: S19390

R:Olliver, S.G.; Anwar, R.; Brown, A.; Gent, M.E.; Indge, K.J.; James, C.M.; Stateva,
submitted to the Protein Sequence Database, March 1992

A:Reference number: S19390

A:Accession: S19390

A:Molecule type: DNA

A:Residues: 1-254 <OLIT>

A:Cross-references: EMBL:X59720; NID:q1907116; PID:q1907139; GSPDB:GN00003; MIPS:YCL0

C:Genetics:

A:Gene: MIPS:YCL005w

A:Map position: 3L

C:Superfamily: Saccharomyces probable membrane protein YCL005w

C:Keywords: transmembrane protein

F:14-31/Domain: transmembrane #status predicted <TM1>

F:37-53/Domain: transmembrane #status predicted <TM2>

F:62-105/Domain: transmembrane #status predicted <TM3>

Query Match 39.6%; Score 42; DB 2; Length 254;
Best Local Similarity 57.1%; Pred. No. 39;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 5 TALDDGVTVDFDRS 18
| | | | | | | | | |
Db 200 TTTDEPTDIDRS 213

RESULT 15

S52775

hypothetical protein 2 - Chloroflexus aurantiacus

C:Species: Chloroflexus aurantiacus

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C:Accession: S52775

R:Niedermeyer, G.; Shiozawa, J.A.; Lottspeich, F.; Felck, R.G.
FEBS Lett. 342, 61-65, 1994

A:Title: The primary structure of two chlorosome proteins from Chloroflexus aurantiac

A:Reference number: S43678; MUID:94192803

A:Accession: S52775

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-305 <NIE>

A:Cross-references: EMBL:Z34000; NID:9496485; PIDN:CA83969.1; PID:9496488

A:Note: only a part of the coding sequence is given in this paper

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1994

C:Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ0279

Query Match 39.6%; Score 42; DB 1; Length 305;
Best Local Similarity 43.8%; Pred. No. 47;
Matches 7; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

OY 2 PTGTALDDGVTVDFDR 17
| | | | | | | | | |
Db 66 PLGTGFSQSINDYFDR 81

Search completed: March 6, 2001, 12:50:52
Job time: 173 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 12:54:35 ; Search time 38.83 Seconds

(without alignments)
16.634 Million cell updates/sec

Title: US-09-196-161d-5

Perfect score: 106

Sequence: 1 CPTGALDDGVTVDFDRSAA 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	45.3	923	HEPA_HAEIN	P44781 haemophilus
2	45	42.5	387	GCP_RICR	Q9268 rickettsia
3	44	41.5	424	ASP_ANCCA	Q16937 ancylostoma
4	44	41.5	699	YQOA_CAEEL	Q09299 caenorhabdi
5	43.5	41.0	1592	SORL_CHICK	Q98930 g sortilin-
6	42	39.6	69	CSPA_RHIME	Q92356 rhizobium m
7	42	39.6	254	YCA5_YEAST	P25587 saccharomyc
8	42	39.6	374	GCP_STRCO	Q86793 streptomyc
9	42	39.6	575	TRBM_HUMAN	P07204 homo sapien
10	41.5	39.2	380	PGL2_PENOL	Q9Y883 penicillium
11	41	38.7	378	KIAB_ECOLI	O52328 escherichia
12	41	38.7	529	PUR9_ECOLI	P15639 e bifunctio
13	41	38.7	764	HTR2_HAISA	P71410 halobacteri
14	41	38.7	1115	DP3A_BACSU	O34623 bacillus su
15	40.5	38.2	356	GP44_HUMAN	O75487 homo sapien
16	40.5	38.2	713	TS44_GIALA	P21849 giardia lam
17	40	37.7	63	IATP_PICJA	P09940 pichia jadi
18	40	37.7	223	NRFC_ECOLI	P22708 escherichia
19	40	37.7	311	CC23_TRYBB	P54666 trypanosoma
20	40	37.7	530	Y119_NPYAC	P41672 autographa
21	40	37.7	798	T2D4_YEAST	P38129 saccharomyc
22	40	37.7	967	HEPA_ECOLI	P23852 escherichia
23	40	37.7	1198	TP2M_CAEEL	P24534 caenorhabdi
24	40	37.7	1607	LMG1_MOUSE	P24548 mus musculu
25	39.5	37.3	446	TBG_SCHJP	Q9Y882 schistosach
26	39	36.8	164	YK12_MYCTU	Q10845 mycobacteri
27	39	36.8	225	NRFC_HAEIN	P45015 haemophilus
28	39	36.8	337	GCP_ECOLI	P05852 escherichia
29	39	36.8	338	SP54_BACSU	P40869 bacillus su
30	39	36.8	342	GCP_HAEIN	P43764 haemophilus
31	39	36.8	398	GDN_HUMAN	P07093 homo sapien
32	39	36.8	410	NEUS_CHICK	O90935 gallus gall
33	39	36.8	426	TWIN_DROME	Q03019 drosophila

34	39	36.8	428	LE22_METTH	O27668 methanobact
35	39	36.8	630	YD13_SCHPO	Q92341 schizosacch
36	39	36.8	854	WUT5_SALRY	P10339 salmonella
37	39	36.8	1210	EGFR_MOUSE	O01279 mus musculu
38	39	36.8	1238	YNI3_YEAST	P53840 saccharomyc
39	39	36.8	2211	FAS_BOVIN	Q28107 bos taurus
40	38.5	36.3	650	EM86_BOOMI	P20736 boophilus m
41	38.5	36.3	2214	SORL_HUMAN	Q92673 h sortilin-
42	38	35.8	122	RL14_RICR	Q92673 h sortilin-
43	38	35.8	158	RL11_HAIVO	P41200 halobacteri
44	38	35.8	171	ATPF_HELPJ	Q92K77 helicobacte
45	38	35.8	171	ATPF_HELPJ	P56086 helicobacte

ALIGNMENTS

```

RESULT 1
HEPA_HAEIN STANDARD: PRT: 923 AA.
AC P44781;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE RNA POLYMERASE ASSOCIATED PROTEIN HOMOLOG (ATP-DEPENDENT HELICASE
HEPA).
GN HEPA OR HI0616.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Georgagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -1- SUBUNIT: BINDS TO THE RNA POLYMERASE (RNAP) (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL: U32744; AAC22275.1; -.
CC TIGR: HI0616; -.
DR INTERPRO: IPR000330; -.
DR INTERPRO: IPR001650; -.
DR PFAM: PF00176; SNF2_N; 1.
DR PFAM: PF00271; helicase_C; 1.
KW Helicase; ATP-binding.
FT NP_BIND 175 182 ATP (BY SIMILARITY).
FT FT DEAH BOX.
FT SITE 278 281
SQ SEQUENCE 923 AA; 104405 MW; 8022403581DAADBD CRC64;
Query Match 45.3%; Score 48; DB 1; Length 923;
Best Local Similarity 44.4%; Pred. No. 7.8;
Matches 12; Conservative 2; Mismatches 5; Indels 8; Gaps 1;

```

OY 2 PTGTAL-----DDGVTDFVDRSA 20
 11111
 11111
 11111
 Db 697 PTGTALVDFPGLKEGVTVDFDRELA 723

RESULT 2

GCP_RICPR STANDARD; PRT; 387 AA.
 AC Q9ZEA8;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE PROBABLE O-SIALOGLYCOPROTEIN ENDOPEPTIDASE (EC 3.4.24.57)
 DE (GLYCOPROTEASE).
 GN GCP OR RP037.
 OS Rickettsia prowazekii.
 OC Rickettsia prowazekii; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiidae; Rickettsia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MADRID E;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomerodipour A., Andersson J.O.,
 RA Sicheit-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria."
 RL Nature 396:133-140(1998).
 CC -1- FUNCTION: COULD BE A METALLOPROTEASE.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF O-SIALOGLYCOPROTEINS; CLEAVES
 CC 31-ARG-1-ASP-32 BOND IN GLYCOPHORIN A. DOES NOT CLEAVE
 CC UNGLYCOSYLATED PROTEINS, DESIALYLATED GLYCOPROTEINS OR
 CC GLYCOPROTEINS THAT ARE ONLY N-GLYCOSYLATED.
 CC -1- COFACTOR: ZINC (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M22; ALSO KNOWN AS THE
 CC GLYCOPROTEASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ235270; CAI4508.1; -
 CC DR INTERPRO: IPR000905; -
 CC DR PRAM: PF00814; Peptidase_M22; 1.
 CC DR PRINTS: PR00789; OSIALOPTASE.
 CC DR PROSITE: PS01016; GLYCOPROTEASE; 1.
 CC KW Hydrolase; Metalloprotease; Zinc.
 CC FT METAL 112 112 ZINC (POTENTIAL).
 CC FT METAL 116 116 ZINC (POTENTIAL).
 CC SQ SEQUENCE 387 AA; 42725 MW; AB974E50F138591D CRC64;

Query Match 42.5%; Score 45; DB 1; Length 387;
 Best Local Similarity 50.0%; Pred. No. 9.7;
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 4 GTALDDGVTDFVDRSA 19
 11111
 11111
 11111
 Db 155 GTTIDAVGTFDKVA 170

RESULT 3

ASP_ANCCA STANDARD; PRT; 424 AA.
 AC Q16937;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE ANCYLOSTOMA SECRETED PROTEIN PRECURSOR.

GN ASP.
 OS Ancylostoma caninum (Dog hookworm).
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
 CC Ancylostomatidae; Ancylostomatidae; Ancylostomatinae; Ancylostoma.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96215086; PubMed=8636085;
 RA Hawdon J.M., Jones B.F., Hoffman D.R., Hotez P.J.;
 RT "Cloning and characterization of Ancylostoma-secreted protein. A
 RT novel protein associated with the transition to parasitism by
 RT infective hookworm larvae."
 RL J. Biol. Chem. 271:6672-6678(1996).
 CC -1- FUNCTION: ASSOCIATED WITH THE TRANSITION TO PARASITISM BY
 CC INFECTIVE HOOKWORM LARVAE.
 CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
 CC INSECTS AG3/AG5; FUNGI SCT/SC14 AND PLANTS PR-1.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: U26187; AAC47001.1; -
 DR HSPD: P04284; ICPE.
 DR INTERPRO: IPR001283; -
 DR PRAM: PF00188; SCP; 1.
 DR PRINTS: PR00837; VSTPXLIKE.
 DR PROSITE: PS01009; SCP_AG5_PRI_SCT_1; FALSE_NEG.
 DR PROSITE: PS01010; SCP_AG5_PRI_SCT_2; FALSE_NEG.
 CC KW Signal.
 CC FT SIGNAL 1 18 POTENTIAL.
 CC FT CHAIN 19 424 ANCYLOSTOMA SECRETED PROTEIN.
 CC SQ SEQUENCE 424 AA; 45735 MW; ABC8295F5D3035F9 CRC64;

Query Match 41.5%; Score 44; DB 1; Length 424;
 Best Local Similarity 46.7%; Pred. No. 15;
 Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 1 CPTGALDDGVTDFV 15
 11111
 11111
 11111
 Db 227 CPSNTGMTDSVDRTE 241

RESULT 4

YQOA_CAEEL STANDARD; PRT; 699 AA.
 ID YQOA_CAEEL
 AC Q09299;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHETICAL 76.5 KDA PROTEIN EEBD8.10 IN CHROMOSOME II.
 GN EEBD8.10.
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 CC Rhabditidae; Peloderinae; Caenorhabditis.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Chisoe S.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----


```

DR EMBL: U23484: AAC46770.1: -
DR MORREP: EED8.10: CE01883.
DR INTERPRO: IPR000504: -.
DR PFM: PF00076: trm; 1.
DR PROSITE: PS50181: EBOX; 1.
DR PROSITE: PS50102: RRM; 1.
KW Hypothetical protein.
SQ SEQUENCE 699 AA: 76517 MW: EDLFLID965464F2B9 CRC64:

Query Match 41.5%; Score 44; DB 1; Length 699;
Best Local Similarity 58.8%; Pred. No. 25;
Matches 10; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 4 GIALDDGVTDVDFDRSAA 20
      | | | | | | | | | |
Db 87 GPAFDGGTFDDFDASIA 103

RESULT 5
SORL_CHICK STANDARD: PRG: 1592 AA.
AC Q98930:
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE SORLILIN-RELATED RECEPTOR (SORLING PROTEIN-RELATED RECEPTOR CONTAINING
DE LDLR CLASS A REPEATS) (SORLA-1) (LOW-DENSITY LIPOPROTEIN
DE RECEPTOR RELATIVE WITH 11 LIGAND-BINDING REPEATS) (LDLR RELATIVE WITH
DE 11 LIGAND-BINDING REPEATS) (LR11) (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauromia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN:
RX MEDLINE=97301565; PubMed=9157966;
RA Morwald S., Yamazaki H., Bujo H., Kusunoki J., Kanaki T., Seimiya K.,
RA Morisaki N., Nimpf J., Schneider W.J., Saito Y.;
RT "A novel mosaic protein containing LDL receptor elements is highly
RT conserved in humans and chickens.";
RL Arterioscler. Thromb. Vasc. Biol. 17:996-1002(1997).
CC -!- FUNCTION: LIKELY TO BE A MULTIFUNCTIONAL ENDOCYTIC RECEPTOR, THAT
CC MAY BE IMPLICATED IN THE UPTAKE OF LIPOPROTEINS AND OF PROTAGES.
CC BINDS LDL, THE MAJOR CHOLESTEROL-CARRYING LIPOPROTEIN OF PLASMA,
CC AND TRANSPORTS IT INTO CELLS BY ENDOCYTOSIS. BINDS THE RECEPTOR-
CC ASSOCIATED PROTEIN (RAP). COULD PLAY A ROLE IN CELL-CELL
CC INTERACTION. BOTH ESTROGEN STATUS AND CHOLESTEROL LEVELS LACK
CC REGULATORY EFFECTS ON THIS RECEPTOR.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
CC -!- TISSUE SPECIFICITY: EXPRESSED ABUNDANTLY IN BRAIN. PRESENT IN THE
CC TESTES, ADRENAL GLANDS AND DETECTABLE IN THE LUNG.
CC -!- SIMILARITY: CONTAINS 5 BNR REPEATS.
CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 11 LDL-RECEPTOR CLASS A DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 FIBROBLAST TYPE III-LIKE DOMAINS.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y08109; CA669324.1; -.
CC HSSP: P01130; IABJ.
CC INTERPRO: IPR000003: -.
CC INTERPRO: IPR0000561: -.
CC INTERPRO: IPR001777: -.
CC INTERPRO: IPR002172: -.
CC INTERPRO: IPR002860: -.

```

[illegible]

```

SQ  SEQUENCE 1592 AA: 178409 MW: 24EDAA5BA21B203 CRC64:
Query Match 41.0%; Score 43.5; DB 1; Length 1592;
Best Local Similarity 39.1%; Pred. No. 66;
Matches 9; Conservative 3; Mismatches 6; Indels 5; Gaps 1.

OY 1 CPTGALD-----DGVTVFDRS 18
      ||||: :||| | |
Db 1150 CPGNCISTSKHNGITDCADAS 1172

RESULT 6
CSPA_RHIME STANDARD: PRT; 69 AA.
AC Q9Z356;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE COLD SHOCK PROTEIN CSPA.
OS CSPA.
SN Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
CC Rhizobiaceae; Sinorhizobium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=20087567; PubMed=10618253;
RA O'Connell K.P., Thomasow M.F.;
RT "Transcriptional organization and regulation of a polycistronic cold
RT shock operon in Sinorhizobium meliloti RM1021 encoding homologs of
RT the Escherichia coli major cold shock gene cspa and ribosomal protein
RT gene rpsu.";
RL Appl. Environ. Microbiol. 66:392-400(2000).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- INDUCTION: BY COLD SHOCK.
CC -1- SIMILARITY: BELONGS TO THE COLD-SHOCK DOMAIN (CSD) FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce
CC or send an email to license@sib-sib.ch).
-----
CC CC
CC EMBL; AF030523; AAC64672.1; -.
DR HSSP; P15277; 3MER.
DR INTERPRO; IPR002059; -.
DR PRAM; PF00313; CSD; 1.
DR PRINTS; PR00050; COLD_SHOCK.
DR PROSITE; PS00352; COLD_SHOCK; 1.
KW Transcription regulation; DNA-binding; Activator.
FT DOMAIN 4 66 CSD.
FT SEQUENCE 69 AA: 7424 MW: 682897E77CD2999E CRC64;

OY 8 DDGVDVDFDRSA 20
      ||| |||| :||
Db 21 DDGADVDFVHSA 33

Query Match 39.6%; Score 42; DB 1; Length 69;
Best Local Similarity 61.5%; Pred. No. 5.2;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0.

RESULT 7
YCAS_YEAST STANDARD: PRT; 254 AA.
AC P25587;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)

```

```

DE HYPOTHEICAL 28.7 KDA PROTEIN IN CWH36-PELI INTERGENIC REGION.
DS Saccharomyces cerevisiae (Baker's yeast).
OS YCLO05W OR YCL5W.
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
RN Saccharomycetaceae; Saccharomycetes.
[1]
RP SEQUENCE FROM N.A.
RA Oliver S.G., Anwar R., Brown A., Gent M.E., Indge K.J., James C.M.,
RA Staveley I.,
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X59720; CAA42354.1; -.
DR PIR; S19390; S19390..
DR SGD; S0000511; YCLO05W.
DR KW Hypothetical protein.
SQ SEQUENCE 254 AA; 28672 MW; 4EB7E7FE073D8A41 CRC64;

Query Match 39.6%; Score 42; DB 1; Length 254;
Best Local Similarity 57.1%; Pred. No. 19;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 5 TALDGVGVDFVDRS 18
| | | | |
Db 200 TTTDDEPTDIWDRS 213

RESULT 8
GCP_STRCO STANDARD: PRT: 374 AA.
ID GCP_STRCO AC 086793;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PROBABLE O-STALOGLYCOPROTEIN ENDOPEPTIDASE (EC 3.4.24.57)
DE (GLYCOPROTEINASE).
GN GCP OR SC6G4.30.
OS Streptomyces coelicolor.
OS Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RA SPRAIN-A3(2);
RA Saunders D., Harris D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
CC -----
CC -1- FUNCTION: COULD BE A METALLOPROTEASE.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF O-STALOGLYCOPROTEINS; CLEAVES
CC 31-ARF-1-ASP-32 BOND IN GLYCOPHORIN A. DOES NOT CLEAVE
CC UNGLYCOSYLATED PROTEINS, DESIALYLATED GLYCOPROTEINS OR
CC GLYCOPROTEINS THAT ARE ONLY N-GLYCOSYLATED.
CC -1- COFACTOR: ZINC (PROBABLE).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M22; ALSO KNOWN AS THE
CC GLYCOPROTEINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL031317; CAA20408.1; -.
DR INTERPRO; IPR000905; -.
DR PFAM; PF00814; Peptidase_M22; 1.

```

DR PRINTS: PR00789; OSTIALOPTASE.
DR PROSITE: PS01016; GLYCOPROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc.
FT METAL 117 117 ZINC (POTENTIAL).
FT METAL 121 121 ZINC (POTENTIAL).
SQ SEQUENCE 374 AA; 38986 MW; 1259FE6DAED0DFE8 CRC64;

Query Match 39.6%; Score 42; DB 1; Length 374;
Best Local Similarity 38.9%; Pred. No. 27;
Matches 7; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 2 PGTALDGCVTVEFDRSA 19
DB 160 PLGATIDDAAGEAFDKIA 177

RESULT 9
TRBM_HUMAN STANDARD; PRT; 575 AA.
ID TRBM_HUMAN
AC P07204;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE THROMBOMODULIN PRECURSOR (FETOMODULIN) (TM) (CD141 ANTIGEN).
GN THBD OR THRM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86004395; PubMed=2820710;
RA Suzuki K., Kusumoto H., Deyashiki Y., Nishioke J., Maruyama I.,
Zushi M., Kawahara S., Honda G., Yamamoto S., Horiguchi S.;
RT "Structure and expression of human thrombomodulin, a thrombin
RT receptor on endothelium acting as a cofactor for protein C
RT activation.";
RL EMBO J. 6:1891-1897(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86024950; PubMed=2822087;
RA Wen D., Dittman W.A., Ye R.D., Deaven L.L., Majerus P.W., Sadler J.E.;
RT "Human thrombomodulin: complete cDNA sequence and chromosome
RT localization of the gene.";
RL Biochemistry 26:4350-4357(1987).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=87317665; PubMed=2819876;
RA Jackson R.W., Beeler D.L., Fritze L., Soff G., Rosenberg R.D.;
RT "Human thrombomodulin gene is intron depleted: nucleic acid sequences
RT of the cDNA and gene predict protein structure and suggest sites of
RT regulatory control.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:6425-6429(1987).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=88227901; PubMed=2836377;
RA Shirai T., Shiojiri S., Ito H., Yamamoto S., Kusumoto H.,
RA Deyashiki Y., Maruyama I., Suzuki K.;
RT "Gene structure of human thrombomodulin, a cofactor for thrombin-
RT catalyzed activation of protein C.";
RL J. Biochem. 103:281-285(1988).
RN [5]
RP CARBOHYDRATE-LINKAGE SITE SER-492, AND MUTAGENESIS.
RX MEDLINE=94029900; PubMed=8216207;
RA Gerlitz B., Hassell T., Vlahos C.J., Parkinson J.F., Bang N.U.,
RA Grinnell B.W.;
RT "Identification of the predominant glycosaminoglycan attachment site
RT in soluble recombinant human thrombomodulin: potential regulation of
RT functionality by glycosyltransferase competition for serine474.";
RL Biochem. J. 295:131-140(1993).
RN [6]
RP STRUCTURE BY NMR OF 389-407.
RX MEDLINE=96007474; PubMed=7559494;

RA Adler M., Seto M.H., Nitecki D.E., Lin J.H., Light D.R., Morser J.;
RT "The structure of a 19-residue fragment from the C-loop of the fourth
RT epidermal growth factor-like domain of thrombomodulin.";
RL J. Biol. Chem. 270:23366-23372(1995).
RN [7]
RP STRUCTURE BY NMR OF 364-407.
RX MEDLINE=96100636; PubMed=8528067;
RA Meinelinger D.P., Hunter M.J., Komives E.A.;
RT "Synthesis, activity, and preliminary structure of the fourth
RT EGF-like domain of thrombomodulin.";
RL Protein Sci. 4:1683-1695(1995).
RN [8]
RP STRUCTURE BY NMR OF 427-444.
RX MEDLINE=95034791; PubMed=7947766;
RA Srinivasan J., Hu S., Hrabal R., Zhu Y., Komives E.A., Ni F.;
RT "Thrombin-bound structure of an EGF subdomain from human
RT thrombomodulin determined by transferred nuclear Overhauser
RT effects.";
RL Biochemistry 33:13553-13560(1994).
RN [9]
RP STRUCTURE BY NMR OF 427-444.
RX MEDLINE=96676211; PubMed=8745396;
RA Hrabal R., Komives E.A., Ni F.;
RT "Structural resiliency of an EGF-like subdomain bound to its target
RT protein, thrombin.";
RL Protein Sci. 5:195-203(1996).
RN [10]
RP STRUCTURE BY NMR OF 405-444.
RX MEDLINE=96035729; PubMed=9367781;
RA Sampoll Benitez B.A., Hunter M.J., Meinelinger D.P., Komives E.A.;
RT "Structure of the fifth EGF-like domain of thrombomodulin: an
RT EGF-like domain with a novel disulfide-bonding pattern.";
RL J. Mol. Biol. 273:913-926(1997).
CC -FUNCTION: THROMBOMODULIN IS A SPECIFIC ENDOTHELIAL CELL RECEPTOR
CC THAT FORMS A 1:1 STOICHIOMETRIC COMPLEX WITH THROMBIN. THIS
CC COMPLEX IS RESPONSIBLE FOR THE CONVERSION OF PROTEIN C TO THE
CC ACTIVATED PROTEIN C (PROTEIN CA). ONCE EVOLVED, PROTEIN CA
CC SCISSIONS THE ACTIVATED COFACTORS OF THE COAGULATION MECHANISM,
CC FACTOR VA AND FACTOR VIIIA, AND THEREBY REDUCES THE AMOUNT OF
CC THROMBIN GENERATED.
CC -SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -TISSUE SPECIFICITY: ENDOTHELIAL CELLS ARE UNIQUE IN SYNTHESIZING
CC THROMBOMODULIN.
CC -SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC -DATABASE: NAME=PROM; NOTE=CD guide CD141 entry.
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd141.htm".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X05495; CAA29045.1; -
CC EMBL: M16552; AAB59508.1; -
CC EMBL: J02973; AAB61175.1; -
CC EMBL: D00210; BAA00149.1; -
CC PIR: A27073; A27073.
CC PIR: A28307; A28307.
CC PIR: A29680; A29680.
CC PDB: 1EGT; 15-NOV-95.
CC PDB: 1EGD; 20-JUN-96.
CC PDB: 1FGE; 20-JUN-96.
CC PDB: 1TMR; 08-JUN-95.
CC PDB: 1Z4Q; 29-JAN-96.
CC PDB: 1ADX; 24-DEC-97.
CC PDB: 2ADX; 24-DEC-97.
CC MIM: 188040; -
CC INTERPRO: IPR000152; -
CC INTERPRO: IPR000561; -
CC INTERPRO: IPR001304; -

```

DR INTERPRO: IPR001491; -
DR INTERPRO: IPR001881; -
DR PFAM: PF00008; EGF; 5.
DR PFAM: PF00059; lectin_c; 1.
DR PRINTS: PR00907; THROMBOMODULIN.
DR PROSITE: PS00010; ASX HYDROXYL; 2.
DR PROSITE: PS00022; EGF_1; FALSE_NEG.
DR PROSITE: PS01186; EGF_2; 2.
DR PROSITE: PS01187; EGF_CA; 2.
DR PROSITE: PS50041; C-TYPE LECTIN_2; 1.
DR PROSITE: PS50041; Receptor; Blood coagulation; Repeat; Transmembrane;
KW Endothelial cell; Receptor; Blood coagulation; Repeat; Transmembrane;
KW Glycoprotein; Signal; EGF-like domain; 3D-structure.
FT SIGNAL 1 21
FT CHAIN 22 575
FT DOMAIN 22 515
FT TRANSSEM 516 539
FT DOMAIN 540 575
FT DOMAIN 241 281
FT DOMAIN 284 324
FT DOMAIN 325 363
FT DOMAIN 365 405
FT DOMAIN 404 440
FT DOMAIN 441 481
FT DISULFID 245 256
FT DISULFID 252 265
FT DISULFID 267 280
FT DISULFID 288 296
FT DISULFID 292 308
FT DISULFID 310 323
FT DISULFID 329 340
FT DISULFID 336 349
FT DISULFID 351 362
FT DISULFID 369 378
FT DISULFID 374 388
FT DISULFID 390 404
FT DISULFID 408 413
FT DISULFID 417 425
FT DISULFID 427 439
FT DISULFID 445 455
FT DISULFID 451 464
FT DISULFID 466 480
FT CARBOHYD 47 47
FT CARBOHYD 115 115
FT CARBOHYD 116 116
FT CARBOHYD 174 174
FT CARBOHYD 334 334
FT CARBOHYD 382 382
FT CARBOHYD 409 409
FT CARBOHYD 411 411
FT CARBOHYD 492 492
FT CARBOHYD 498 498
FT CARBOHYD 504 504
FT CONFLICT 473 473
SQ SEQUENCE 575 AA; 60329 MW; 9AF03CD151227D52 CRC64;

Query Match 39.6%; Score 42; DB 1; Length 575;
Best Local Similarity 56.2%; Pred. No. 42;
Matches 9; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

```

```

CN PG2.
OS Penicillium olsonii.
OS Eukaryota; Fungi; Ascomycota; amorphous Ascomycota; Penicillium.
CC [1]
RN SEQUENCE FROM N.A.
RA Wagner F.;
RL Thesis (1999), University of Hamburg, Germany.
CC -1- CATALYTIC ACTIVITY: RANDOM HYDROLASES OF 1,4-ALPHA-D-
CC GALACTOSIDURONIC LINKAGES IN PECTATE AND OTHER GALACTURONANS.
CC -1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES
CC (POLYGALACTURONASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ243522; CAB46909.1; -
DR INTERPRO: IPR000743; -
DR PFAM: PF00295; Glyco hydro. 28; 1.
DR PROSITE: PS00502; POLYGALACTURONASE; 1.
KW Hydrolyase; Glycosidase; Cell wall; Signal.
FT SIGNAL 1 20
FT PROPEP 21 38
FT CHAIN 39 380
FT CARBOHYD 287 287
SQ SEQUENCE 380 AA; 38652 MW; 2B719D8DC4E23A46 CRC64;

Query Match 39.2%; Score 41.5; DB 1; Length 380;
Best Local Similarity 45.0%; Pred. No. 33;
Matches 9; Conservative 3; Mismatches 3; Indels 5; Gaps 1;

```

```

OY 1 PG2ALDDG--VTDV 14
Db 427 CPFGYILDDGFCITDI 442

RESULT 10
PG12_PENOL STANDARD; PRT; 380 AA.
AC OY833;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE POLYGALACTURONASE 2 PRECURSOR (EC 3.2.1.15) (PG 2) (PECTINASE 2).

```

```

OY 2 PG2ALD-----DGVTDVFD 16
Db 71 PG2TLDLTGLDGTTFIVE 90

RESULT 11
KLAB_ECOLI STANDARD; PRT; 378 AA.
AC Q52328; Q52326;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE KLAB PROTEIN (TELA PROTEIN).
GN KLAB OR TELA.
OS Escherichia coli.
OC Plasmid Incp-alpha RK2.
CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.
CC [1]
RN SEQUENCE FROM N.A.
RA MEDLINE=91258329; PubMed=2045366;
RA Goncharoff P., Saadi S., Chang C.H., Saitman L.H., Figurski D.H.;
RT "Structural, molecular, and genetic analysis of the kila operon of
RT broad-host-range plasmid RK2."
RL J. Bacteriol. 173:3463-3477(1991).
RN [2]
RN SEQUENCE FROM N.A.
RA MEDLINE=9123183; PubMed=1846856;
RA Walter E.G., Thomas C.M., Ibbotson J.P., Taylor D.E.;
RT "Transcriptional analysis, translational analysis, and sequence of
RT the kila-tellurite resistance region of plasmid RK2ter."
RL J. Bacteriol. 173:1111-1119(1991).
CC -1- FUNCTION: BELONGS TO THE KILA OPERON, WHICH IS ASSOCIATED WITH
CC CRYPTIC TELLURITE RESISTANCE, AND INCM PLASMID FERTILITY
CC INHIBITION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M62426; AAA26402.1; -;
 DR EMBL; M62846; AAA98154.1; -;
 DR Plasmid; Tellurium resistance.
 SQ SEQUENCE 378 AA; 42156 MW; 8297D18679DA3672 CRC64;
 Query Match 38.7%; Score 41; DB 1; Length 378;
 Best Local Similarity 50.0%; Pred. No. 40;
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 QY 5 TALDDGVTDFDRSA 20
 DB 283 TAIIDDTMDLKRMA 298
 RESULT 12
 PUR9_ECOLI STANDARD; PRT; 529 AA.
 AC P15639;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE BIFUNCTIONAL PURINE BIOSYNTHESIS PROTEIN PURH [INCLUDES:
 DE PHOSPHORIBOSYLAMINOIMIDAZOLECARBOXAMIDE FORMYLTRANSFERASE (EC 2.1.2.3)
 DE (ATICAR TRANSFORMYLASE); IMP CYCLOHYDROLASE (EC 3.5.4.10) (IMOSINICASE)
 DE (IMP SYNTHETASE) (ATIC)].
 GN PURH.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 CC [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-90078227; PubMed-2687276;
 RA Alba A., Mizobuchi K.;
 RT "Nucleotide sequence analysis of genes purH and purD involved in the
 RT de novo purine nucleotide biosynthesis of Escherichia coli.";
 RL J. Biol. Chem. 264:21239-21246(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN-K12;
 RC MEDLINE-90286915; PubMed-2192230;
 RA Flannigan K.A., Hennigan S.H., Vogelbacker H.H., Gots J.S.,
 RA Smith J.M.;
 RT "Purine biosynthesis in Escherichia coli K12: structure and DNA
 RT sequence studies of the purH locus.";
 RL Mol. Microbiol. 4:381-392(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN-K12 / MG1655;
 RC MEDLINE-94089392; PubMed-265357;
 RA Blatter F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
 RA Daniels D.L.;
 RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the
 RT region from 89.2 to 92.8 minutes.";
 RL Nucleic Acids Res. 21:5408-5417(1993).
 RN [4]
 RP SEQUENCE OF 1-12;
 RX STRAIN-K12 / EMG2;
 RC MEDLINE-97443975; PubMed-9298646;
 RA Link A.J., Robison K., Church G.M.;
 RT "Comparing the predicted and observed properties of proteins encoded
 RT in the genome of Escherichia coli K-12.";
 RL Electrophoresis 18:1259-1313(1997).
 CC -1- CATALYTIC ACTIVITY: 10-FORMYLTETRAHYDROFOLATE + 5'-PHOSPHORIBOSYL-
 CC 5-AMINO-4-IMIDAZOLECARBOXAMIDE = TETRAHYDROFOLATE +
 CC 5'-PHOSPHORIBOSYL-5-FORMAMIDO-4-IMIDAZOLECARBOXAMIDE.

CC -1- CATALYTIC ACTIVITY: IMP + H(2)O = 5-FORMAMIDO-1-(5-PHOSPHORIBOSYL)
 CC IMIDAZOLE-4-CARBOXAMIDE.
 CC -1- PATHWAY: NINTH AND TENTH STEP IN DE NOVO PURINE BIOSYNTHESIS.
 CC -1- DOMAIN: THE IMP CYCLOHYDROLASE ACTIVITY RESIDES IN THE N-TERMINAL
 CC REGION (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE PURH FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; J05126; AAA4454.1; -;
 DR EMBL; X51950; CAA36212.1; -;
 DR EMBL; U00006; AAC43104.1; -;
 DR EMBL; AE000473; AAC76980.1; -;
 DR PIR; A34193; DTECPH.
 DR ECOGENE; EG10795; PURH.
 DR INTERPRO; IPR002695; -;
 DR PFAM; PF01808; AICARFT_IMPHas; 1
 DR Purine biosynthesis; Transferase; Hydrolase; Multifunctional enzyme.
 SQ SEQUENCE 529 AA; 57329 MW; DC034ED01915DA68 CRC64;
 Query Match 38.7%; Score 41; DB 1; Length 529;
 Best Local Similarity 35.3%; Pred. No. 55;
 Matches 6; Conservative 5; Mismatches 6; Indels 0; Gaps 0;
 QY 2 PTGTALDDGVTDFDRS 18
 DB 291 PCGVAIGNSILDAYRA 307
 RESULT 13
 HTR2_HALSA STANDARD; PRT; 764 AA.
 ID HTR2_HALSA
 AC P71410;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE SENSOR RHODOPSIN II TRANSDUCER (HTR-II) (METHYL-ACCEPTING PHOTOTAXIS
 DE PROTEIN II) (MP-II).
 GN HTRII.
 OS Halobacterium salinarum.
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
 CC Halobacterium.
 CC [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=FLX15;
 RC MEDLINE-96323203; PubMed-8710852;
 RA Zhang W., Brooun A., Mueller M.W., Alam M.;
 RT "The primary structures of the Archaeon Halobacterium salinarum blue
 RT light receptor sensory rhodopsin II and its transducer, a methyl-
 RT accepting protein.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:8230-8235(1996).
 CC -1- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR
 CC SENSOR RHODOPSIN II (SR-II) TO THE FLAGELLAR MOTOR. RESPONDS TO
 CC LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION.
 CC -1- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U62676; AAC44369.1; -;
 DR INTERPRO; IPR001122; -;

Search completed: March 6, 2001, 12:54:37
Job time: 394 sec

THIS PAGE BLANK (USTO)